The Dissertation Committee for Pushkar Kumar Jain
certifies that this is the approved version of the following dissertation:

Dynamically Adaptive Data-driven Simulation of Extreme Hydrological Flows

Committee:

Clint Dawson, Supervisor

Thomas J. R. Hughes

Omar Ghattas

Tan Bui-Thanh

Dynamically Adaptive Data-driven Simulation of Extreme Hydrological Flows

by

Pushkar Kumar Jain

DISSERTATION

Presented to the Faculty of the Graduate School of
The University of Texas at Austin
in Partial Fulfillment
of the Requirements
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN
August 2018

Dedicated to my parents and my sisters.

Acknowledgments

I wish to thank the multitudes of people who helped me. First and foremost, I thank my parents and my sisters for all the support and being beside me throughout the journey.

I would like to express my sincere gratitude to my advisor, Dr. Clint Dawson who inspired me to pursue a doctoral degree. His continuous guidance and support in my Ph.D helped me throughout my research. I am glad that I got an opportunity to work for him.

I would like to thank my thesis committee: Dr. Thomas J. R. Hughes, Dr. Omar Ghattas, and Dr. Tan Bui-Thanh, for their insightful comments. Their expertise in various domains helped me tackle my research problem from various perspectives.

I am thankful to my research colleague Dr. Kyle Mandli for helping me throughout the research topic.

My first professional experience in a research environment that shaped me for the future projects is owed to my Masters supervisor Dr. Corey Trahan.

I am thankful to all the support given by my research group. No words can explain their immense contribution in professional and social arenas. More than being research colleagues, they are my friends. Each one influenced and shaped me in a unique way. In particular I am thankful to Ali, Gajanan, Chen, Deepesh and

Yuxiang for working together on various projects and for all the fun we had in the last few years.

I acknowledge the support of the King Abdullah University of Science and Technology Competitive Research Grant Program, award number OCRF-2014-CRG3-62140389/ORS#2156. I am thankful to Dr. Ibrahim Hoteit and Dr. Omar Knio for providing great insights and support in the research topic.

Last but not the least, I would like to thank my friends Reshma, Amit, Anjali, Akshay, Shreya, Vasudha, Sadhika and Adriano for providing emotional support, stimulating discussions and all-important coffee breaks.

Dynamically Adaptive Data-driven Simulation of Extreme Hydrological Flows

Publication No. __________________

Pushkar Kumar Jain, Ph.D.
The University of Texas at Austin, 2018

Supervisor: Clint Dawson

Hydrological hazards such as storm surges, tsunamis, and rainfall-induced flooding are physically complex events that are costly in loss of human life and economic productivity. Many such disasters could be mitigated through improved emergency evacuation in real-time and through the development of resilient infrastructure based on knowledge of how systems respond to extreme events. Data-driven computational modeling is a critical technology underpinning these efforts. This investigation focuses on the novel combination of methodologies in forward simulation and data assimilation. The forward geophysical model utilizes adaptive mesh refinement (AMR), a process by which a computational mesh can adapt in time and space based on the current state of a simulation. The forward solution is combined with ensemble based data assimilation methods, whereby observations from an event are assimilated into the forward simulation to improve the veracity of the solution, or used to invert for uncertain physical parameters. The novelty in our approach is the tight two-way coupling of AMR and ensemble filtering techniques.

The data assimilation system is implemented on various test cases that delve into the aspects of ensemble based assimilation filters. Additionally, data assimilation on tsunami models is analyzed and a methodology to map the uncertainties in the seabed deformation due to the associated earthquake to the water surface elevation forecast has been presented. Further, using other simulated environments such as the Chile tsunami event of February 2010, a systematic way to calibrate the assimilation system is presented. Finally, the technology is tested by assimilating actual gauge data from the Tohoku tsunami event. These advances offer the promise of significantly transforming data-driven, real-time modeling of hydrological hazards, with potentially broader applications in other science domains.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Background

Coastal areas are often densely populated [12] and are most vulnerable to extreme hazards such as storm surges, hurricanes and tsunamis. As of the 2010 census, 39% of the US population or 123.3 million people are concentrated in areas directly on the shoreline [1]. Furthermore, the world coastal population is projected to rise to around 949 million people (then projected global population is 8.7 billion) by 2030, and to 1.4 billion people (then projected global population is 11.3 billion) by 2060 [83]. This suggests that a large portion of the global population is vulnerable to the destruction that may be caused by these extreme events.

The effects of hazardous flows can be mitigated by accurately forecasting such events and providing timely evacuation measures. Forecasting is done using reliable computational fluid dynamics models that reveal possible inundation regions near the coast as well as other effects of these hazardous flows. Thus, information disseminated from these forecasts can aid in preparedness and undertaking timely preemptive measures when necessary.

This suggests that improved forecasting techniques can prove to be extremely beneficial. In this research, we investigate computational algorithms to advance the science of real-time prediction of coastal and overland hydrological hazards.

1.2 Motivation

The numerical models that simulate large scale geophysical flows involve solving simultaneous partial differential equations that govern the flow of fluids. Under certain assumptions, the governing equations can be used to simulate flows on a synoptic scale, replicating the path and movement of extreme hydrological flows such as hurricanes and tsunamis [68, 56, 25, 70]. The input to these models are in the form of initial condition of the flow, boundary conditions, forcing parameters such as wind stresses, topography of sea bed, salinity etc. Simulations of hydrological hazards are computationally challenging in at least two aspects: they are inherently multi-scale (both spatially and temporally) [65, 2], and there are significant uncertainties in the underlying physical models [46, 3, 20]. These uncertainties are inevitable. This means that uncertainty in any of the inputs can lead to solutions that may not replicate the reality perfectly. For instance, in tsunami modeling, the sea floor motion that generates the tsunami is not well determined because the fault displacement due to the associated earthquake is dependent on the subsurface geologic structure which is seldom known perfectly. A good measure to assess the performance of the models is through validation with field data. This can be in the form of satellite imagery or data from buoys in oceans. However, measurements are not only scarce, but also carry some uncertainties due to various factors such as sensor precision or observation location. For instance, Figure 1.1

shows the locations of Deep-Ocean Assessment and Reporting of Tsunami (DART) sensors placed around the globe to ensure early detection of tsunamis. The sensors are not only sparse, but also the recording of data does not occur continuously in time. Furthermore, effective transmission of real-time information is often not smooth and may carry errors [74].

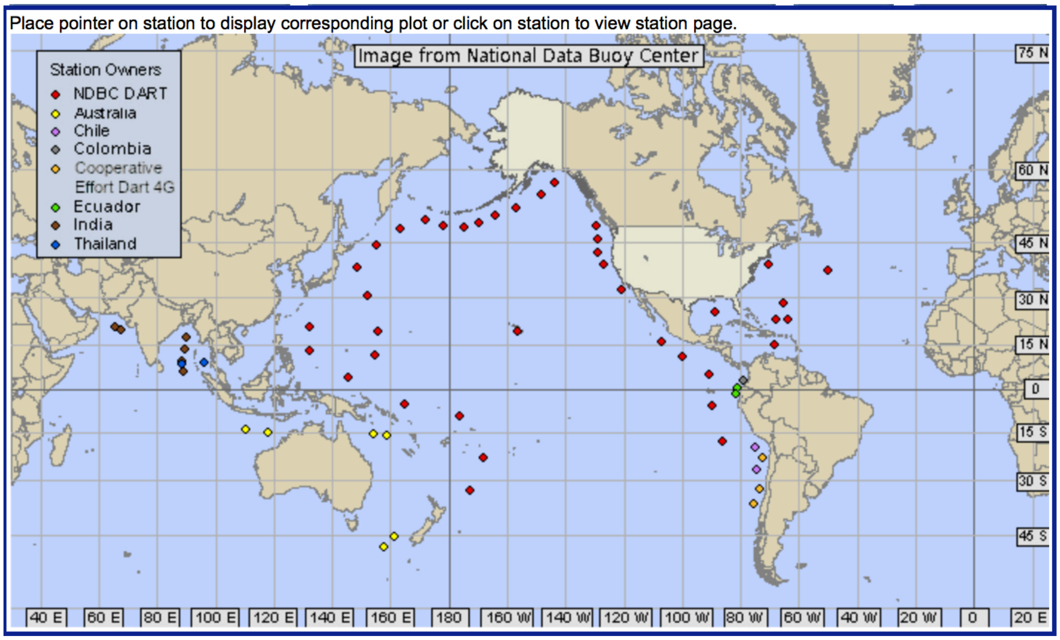


Figure 1.1: DART stations operated as per National Data Buoy Center

In summary, there exists uncertainties in the numerical models as well as the real-time sensor data that are inevitable. The current research uses the technique of data assimilation to improve the forecasting by combining the numerical models and the sensor measurements, taking into the account the respective uncertainties.

1.3 Outline

Ensemble based Kalman filters are a class of data assimilation techniques widely adopted in weather prediction and simulations of large scale geophysical flows. A data assimilation system is a combination of two components - the forecasting model and the assimilation technique. The objective of the research is to develop a data assimilation system using a numerical model, GEOCLAW. Chapter 3 discusses one component of the data assimilation system, that is the assimilation techniques. The chapter describes the traditional Kalman filter, the development of ensemble based techniques and subsequently addresses the state of the art ensemble based assimilation methods. Chapter 4 discusses the second component of the data assimilation system, that is, the forecasting model - GEOCLAW. One of the characteristics of GEOCLAW is the use of a patch based adaptive mesh refinement (AMR) technique to dynamically obtain high resolution solution at the areas of interest. This will be addressed in detail. The current research has been extensively done with respect to tsunami models. Therefore, the numerical model also includes a brief discussion on the parametric model that is used to simulate the onset of the earthquake prior to tsunami and the bathymetry changes that occur subsequently.

Chapter 5 addresses the development of a fully coupled ensemble based assimilation technique with AMR capable numerical model, GEOCLAW. Coupling ensemble based assimilation with AMR can be challenging as the AMR algorithm may dictate the ensemble members to adapt in their own way. We discuss the hurdles faced in development of such a system and devise a methodology to tackle the same. We propose a methodology wherein data are assimilated at all the multiscale levels. Further, while applying the data assimilation framework on tsunami models, we demonstrate an algorithm for the initial ensemble generation, which is essential for any ensemble based filtering technique.

Chapter 4 discusses the numerical results on implementing the coupled system. Specifically, three cases are discussed - 1. *Radial bowl* test case that simulates flow of water in a parabolic bowl with an initial condition given by a Gaussian hump, solely forced by gravity, 2. *Chile tsunami* test case that is a simulation of a tsunami event off the coast of Chile. The test case highlights calibration of a model dealing with a real-world problem. Successful implementation of a data assimilation system involves tuning the hyper-parameters that have been introduced during the development of the assimilation technique; 3. *Tohoku tsunami* test case is another tsunami event off the coast of Japan. We use the lessons learned from *Chile tsunami* test case and implement them on *Tohoku tsunami* event.

Finally, Chapter 7 summarizes the contributions and Chapter 8 provides the conclusions of the study and delves into possible research extensions.

Chapter 2

Literature review

Data assimilation is a technique whereby the observations or sensor measurements along with the uncertainties are combined with the forward models in a statistical way to compute the best possible estimate of the state system. There are two main approaches to assimilating data into numerical models - 1) variational methods such as 3D-Var [26] and 4D-Var [27] and 2) sequential methods [59, 58]. The current research uses the two step sequential approach based on the Kalman filter as proposed by [59, 58] that incorporates data as it is available. The first step is the forecasting stage where the numerical model forecasts the state to a given time. The next step is the analysis stage or the update stage that statistically averages the sensor measurement at that time with the forecasted state according to their estimated uncertainties. Figure 2.1 demonstrates the sequential approach of any data assimilation system such as Kalman filter.

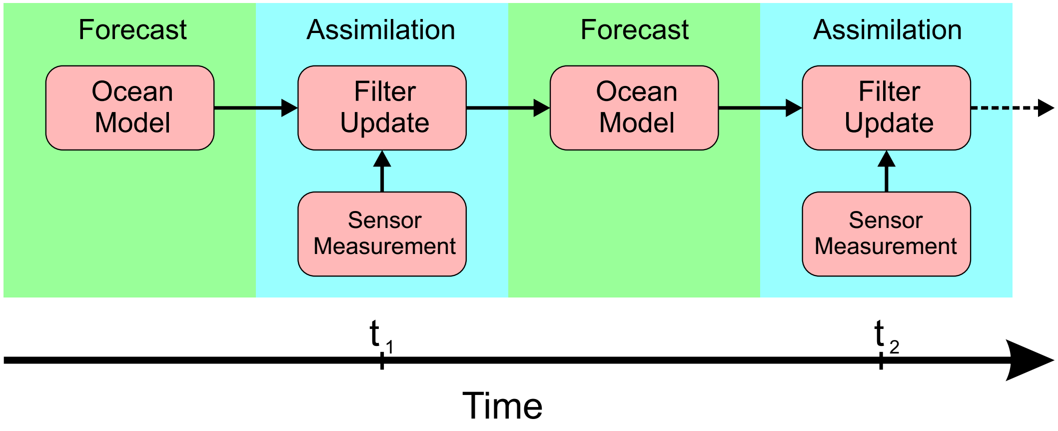


Figure 2.1: Sequential forecast and update steps in Kalman filtering

For linear systems, under the assumption of unbiased model and incoming uncorrelated Gaussian distributed measurement and model errors, the method yields the best linear unbiased estimate. For non-linear dynamics, the Extended Kalman filter was proposed in [91]. These methods, namely Kalman and Extended Kalman filters, propagate the state of the system as well as the associated error covariance, a measure of the uncertainty in the state estimate. Propagation of the error covariance can be computationally prohibitive when the dimension of the state is large. This is typical for large scale hydrodynamic flows, where the dimension of the state can be $\mathcal{O}(10^8)$ [50]. This issue is addressed by ensemble based methods such as the Ensemble Kalman filter [33, 34]. These schemes use Monte Carlo methods to estimate the background error covariance. The method performs an ensemble run of the forward model in the forecast stage from which the sample covariance is calculated and then used at the analysis step. The ensemble based technique asymptotically approaches the Kalman filter in the limit of large ensembles and

Gaussian error distribution [71]. The computational cost has been well described in [34]. Figure 2.2 demonstrates the steps of ensemble based Kalman filters. As seen from the figure, unlike the Kalman filter that performs one forward solve, the ensemble based method runs a an ensemble of forward solves. At the assimilation step, all the members are updated based on the measurement value at the time.

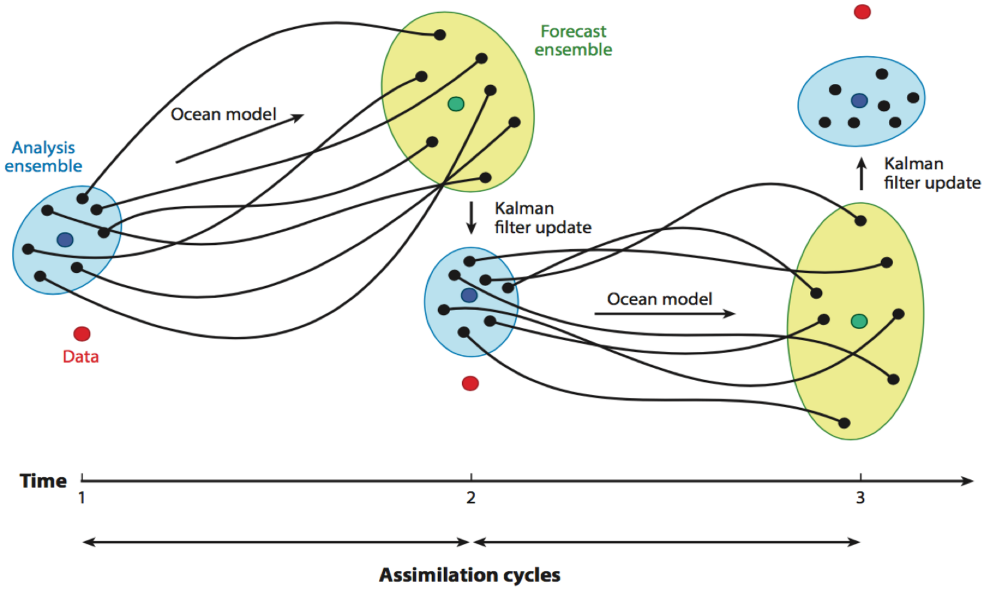


Figure 2.2: Steps of ensemble based Kalman filter. The image has been taken from [31]

Over the years, a variety of ensemble filters have been developed and implemented in various fields [66, 92, 41, 73, 23, 98]. These can be classified into two groups – 1) stochastic, where the analysis states are computed from randomly perturbed observations; and 2) deterministic, where the analysis state is obtained by deterministically transforming the forecast ensemble without perturbing the observations. Several review articles [34, 43, 32, 53, 31, 14] enumerate the advancements in the field, leading to development of various filters, such as optimal interpolation [90], EnSRF [104], EAKF [7, 6], SEIK [86, 51], ETKF [18], SEEK [87, 24], RRSQRT [45], ESTKF [82]. Many of these filters have been developed with the underlying idea of a low-rank approximation of the error covariance matrix.

One of the challenges of data assimilation is that factors such as neglected uncertainties, poorly known model errors and a low rank representation of the estimation error via a finite ensemble size lead to underestimation of the error covariances of the filter. The underestimation of the variance is mitigated by the use of covariance inflation [8, 9, 105, 69, 10, 75]. In addition, localization can be applied to address rank deficiency and spuriously large cross-covariances between different state variables. Two methods of localization often used are 1) Covariance localization [52], where the error covariance matrix is transformed based on physical distance of grid points in the numerical model, and 2) Domain localization [54], where the physical domain is divided into individual local domains for updates. Also, [81] demonstrates a regulated localization scheme that weighs elements of observation error covariance matrix by a localization function of varying width.

2.1 Data assimilation and ocean models

This research aims to develop a data assimilation framework that employs the above described aspects of ensemble based filters with large scale hydrodynamic models, and effectively improve forecasting. Ensemble based filters have been implemented in various large scale oceanic models [60, 61, 35, 67, 23]. [31] reviews the developments in regional data assimilation over the last 15 years.

2.1.1 Tsunami applications

With respect to employing data assimilation with tsunami simulations, significant literature exists. [97] discusses the need and challenges of real time tsunami forecasting with respect to accuracy, speed and robustness of tsunami warning systems. [42] assumes a linear system and assimilates data using optimal interpolation for the 2012 Haida Gwaii earthquake. In this study, the Kalman gain (called the smoothing matrix) does not change with time and is a function of the station distribution. [77, 76] employed the particle filter method to assimilate tide gauge data into a tsunami simulation model. As part of the NOAA tsunami forecasting system [95], the best estimates of tsunami amplitudes and arrival times for potential at-risk communities are obtained by assimilating real-time event data with the Method of Splitting Tsunami (MOST) model [93, 94].

2.2 Data assimilation and adaptive mesh refinement

Some of these large scale ocean models employ adaptive mesh refinement (AMR) technique. Adaptive meshing uses algorithms that dynamically refine the grids spatially, temporally or both, to obtain fine scale solutions in the areas of interest that can be identified based on user criteria [11, 13, 17, 37]. There are two main approaches of implementing AMR. Both the methods identify the regions where fine scale solutions are of interest. One class of AMR technique splits the elements of the grid into finer elements on the same grid [22]. Another class of AMR techniques involve multi-level mesh structure that is comprised of various refinement levels [28, 17]. AMR technique proves to be advantageous towards reducing computational cost. In large scale hydrodynamic models, AMR can be used to dynamically obtain fine scale solutions near the coast, where we are most interested in the solution, while retaining a coarse mesh in deep oceans. Furthermore, in the case of tsunami simulations, we can dynamically refine the mesh along the moving wave front [72, 16].

In the area of data assimilation with AMR, several methods involving 3D-Var and 4D-Var have been proposed by [88, 36]. In a very recent study, [30] implemented the ensemble based filter for data assimilation. The ensembles were interpolated to a reference mesh, while accounting for the interpolation errors using conservative interpolation operators. However, no prior work has implemented ensemble based filter for data assimilation with multi-level AMR, and this research is the first attempt to provide primary insights in this direction. Our goal is to build an advanced data-driven AMR model for efficient simulation and forecasting of coastal marine dynamics.

Chapter 3

Data assimilation

Consider the size of the state to be assimilated is n and the size of the observation vector to be m. Consider a linear dynamical system M_k that forecasts the true state from time t_{k-1} to t_k as,

$$x_k^t = M_k(x_{k-1}^t) + \eta_k,$$

where k is the time step number, $x_k^t \in \mathcal{R}^{n\times 1}$ is the true state (t denotes *truth*) and η_k is the model noise at time step k. The model noise is unknown and assumed to be unbiased with an error covariance of $Q_k \in \mathcal{R}^{n\times n}$. Further, consider the observation or sensor measurements, y_k of the state at time t_k given by,

$$y_k = H_k x_k^t + \epsilon_k,$$

where ϵ_k is the error in the observation operator at time t_k, assumed to be unbiased with error covariance $R_k \in \mathcal{R}^{m\times m}$. The sensor measurements are seldom available throughout the domain. $H_k \in \mathcal{R}^{m\times n}$ is the operator that maps the field of interest from the state space to the observation space.

Hence, we have, $\overline{\eta_k} = 0$, $\overline{\eta_k\eta_k^T} = Q_k$, $\overline{\epsilon_k} = 0$ and $\overline{\epsilon_k\epsilon_k^T} = R_k$.

3.1 Kalman filter

The Kalman filtering is a two-step sequential approach -

1. **Prediction step**: At time t_k, the state forecast, x_k^f is generated from the previous state x_{k-1}^a as,

   $$x_k^f = M_k(x_{k-1}^a),$$

   Note that the superscript a denotes *assimilation* or *update*. Physically, this means that the forward model forecasts the state at the previous time step that is actually the updated step at that time.

   The associated uncertainty in the forecasted state denoted by the forecast error covariance matrix $P_k^f \in \mathcal{R}^{n\times n}$, can be computed as,

   $$\begin{aligned} P_k^f &= \overline{(x_k^t - x_k^f)(x_k^t - x_k^f)^\top} \\ &= M_k P_{k-1}^a M_k^\top + Q_k, \end{aligned} \tag{3.1}$$

2. **Update step**: Given the forecast state is x_k^f and the incoming sensor measurements are y_k, the minimization of the cost function,

   $$\mathcal{J}(x_k) = \frac{1}{2}(x_k^f - x_k)^\top (P_k^f)^{-1}(x_k^f - x_k) + \frac{1}{2}(y_k - H(x_k))^\top R_k^{-1}(y_k - H(x_k)),$$

   gives the assimilated state as,

   $$x_k^a = x_k^f + K_k(y_k - H_k x_k^f), \tag{3.2}$$

   where K_k is Kalman gain matrix at time t_k computed as,

   $$K_k = P_k^f H_k^\top (H_k P_k^f H_k^\top + R_k)^{-1}. \tag{3.3}$$

Physically, the Kalman gain matrix weighs the residual of the observed values and the forecasted state, resulting to the correction term that is added to the forecasted state.

The analysis error covariance is computed by,

$$\begin{aligned} P_k^a &= \overline{(x_k^t - x_k^a)(x_k^t - x_k^a)^\top} \\ &= (\mathbb{I} - K_k H_k) P_k^f (\mathbb{I} - K_k H_k)^\top + (K_k R_k K_k)^\top, \end{aligned} \tag{3.4}$$

where $\mathbb{I}$ is the identity matrix.

Equation (3.4) is also known as *Joseph form* for covariance update. It is valid for any value of K_k. Substituting the optimal Kalman gain in Equation (3.3), the analysis error covariance simplifies to,

$$P_k^a = (\mathbb{I} - K_k H_k) P_k^f$$

Note that the Kalman gain in Equation (3.3) can also be obtained by minimizing the *Joseph* form of analysis error covariance. Hence Kalman filter is also a minimum variance estimate.

Figure 3.1 summarizes the above steps.

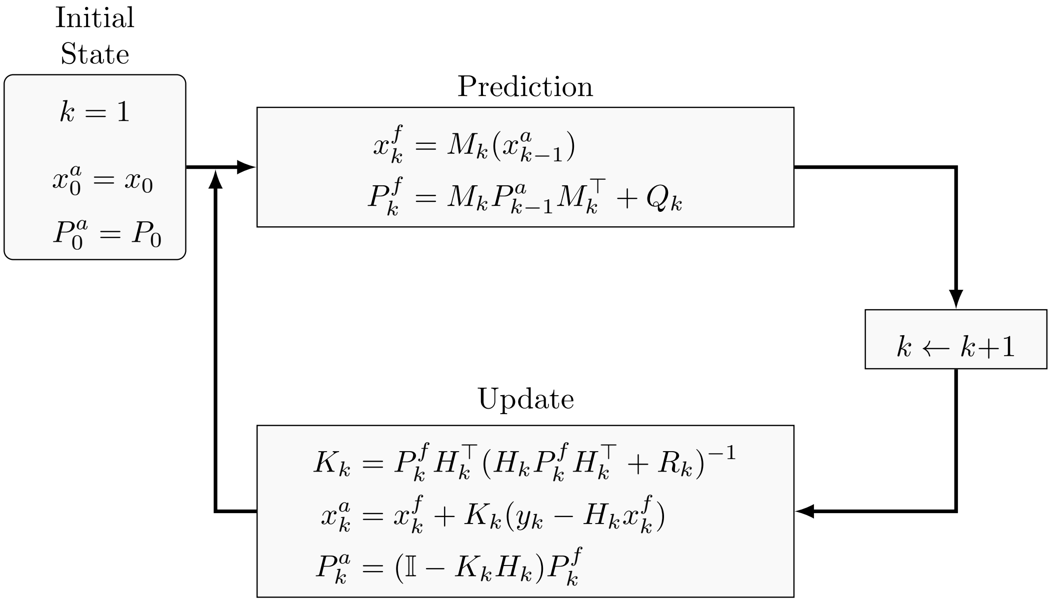


Figure 3.1: Kalman filter algorithm

3.2 Extended Kalman filter (EKF)

Geophysical models typically have non-linear dynamics. Under the assumptions of linear model, the calculation of error covariance matrices simplify and theoretically, the assimilated state is the best linear unbiased estimate. However, the same may not be the case for a non-linear forecasting model.

1. **Prediction step**: For a nonlinear model,$\mathcal{M}_k$, with the forecast equation given by,

   $$x_k^f = \mathcal{M}_k(x_{k-1}^a),$$

   the prediction of error statistics is performed using the Jacobian of $\mathcal{M}$. Hence, if the Jacobian of $\mathcal{M}$ is M, then the predictions steps can be denoted the same way. The forecast error covariance becomes,

$$P_k^f = M_k P_{k-1}^a M_k^\top + Q_k. \tag{3.5}$$

2. **Update step**: The update step remains the same as,Equation (3.2), the Kalman filter in linear model case.

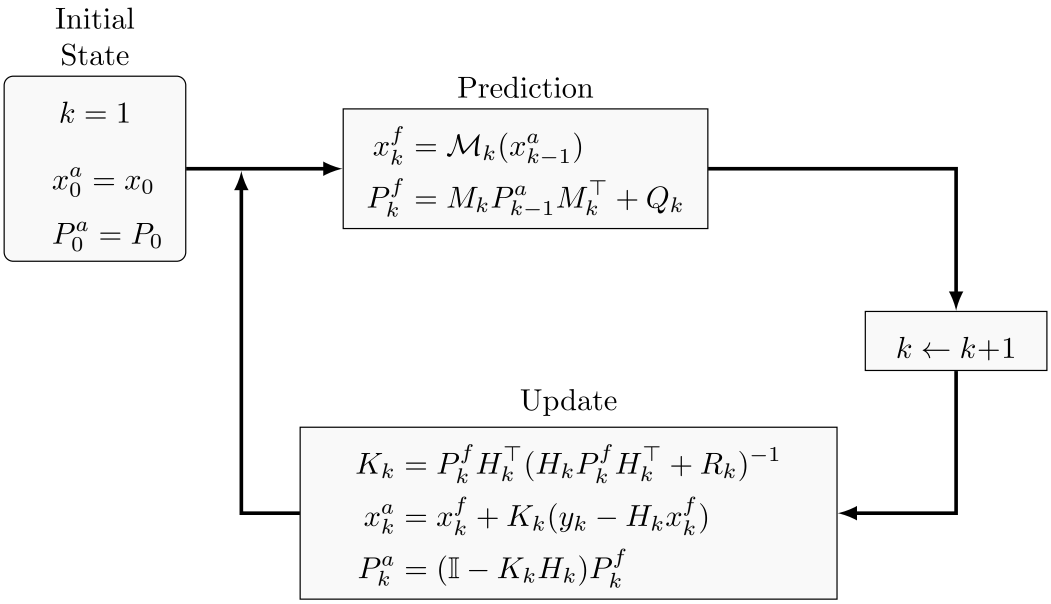


Figure 3.2: Extended Kalman filter algorithm

For large scale geophysical simulations, the dimension of the state vector (x) can be as large as $\mathcal{O}(10^8)$. Hence, the following issues arise -

1. The prediction step that involves propagation of the error covariance matrix (Equation (3.1) or Equation (3.5)) becomes computationally expensive.

2. The size of the error covariance matrices becomes intangible as the storage is $\mathcal{O}(n^2)$.

3. Computation of forward error covariance requires tangent linear operator (Jacobian) of the forward model (M_k).

These bottlenecks motivate us to adopt an ensemble based approach called Ensemble Kalman filter (EnKF).

3.3 Ensemble Kalman filter (EnKF)

The traditional ensemble Kalman filter EnKF described in [33, 34] is formulated as a sequential process of the forecast step followed by the analysis step. Ensemble based Kalman filter is a Monte Carlo approximation to the Kalman filter. Unlike EKF, instead of performing one forward solve, EnKF forecasts an ensemble of N states in the prediction step. These N states can be generated by perturbing the best known state at the initial time.

Using an ensemble of size N, we define the forecasted state matrix $X_k^f \in \mathcal{R}^{n\times N}$ as,

$$X_k^f = \begin{bmatrix} \vdots & & \vdots & & \vdots \\ x_{k,1}^f & \ldots & x_{k,p}^f & \ldots & x_{k,N}^f \\ \vdots & & \vdots & & \vdots \end{bmatrix}, \tag{3.6}$$

where $x_{k,p}^f$ is the p^{th} forecasted ensemble member at time t_k. We also define $\bar{X}_k^f = \bar{x}_k^f \mathbf{1_d}$, where the mean of the forecasted ensemble states is $\bar{x}_k^f = \frac{1}{N}\sum_{i=1}^{i=N} x_{k,i}^f$ and $\mathbf{1_d}$ is the matrix such that all the entries are $1s$. Similar to the forecast variables defined in Equation (3.6), we define X_k^a as the analysis state matrix at time t_k comprising of the analysis state ensemble members $x_{k,p}^a$.

1. **Prediction step**: Using the possibly non-linear dynamic model, $\mathcal{M}$, all the ensemble members are forecasted from time t_{k-1} to t_k using,

   $$X_k^f = \mathcal{M}_k(X_{k-1}^a), \tag{3.7}$$

   At time t_k, in place of using Equation (3.5), the forecast error covariance $P_k^f \in \mathcal{R}^{n \times n}$ is computed as,

   $$P_k^f = \frac{1}{N-1}(X_k^f - \bar{X}_k^f)(X_k^f - \bar{X}_k^f)^\top, \tag{3.8}$$

2. **Update step**: Analogous to the fact that the forecast error covariance in Equation (3.8) is an ensemble approximation to Equation (3.5), an ensemble approximation to the Kalman gain in Equation (3.3) is computed from the same equation, with the forecast error covariance computed from Equation (3.8).

   $$K_k = P_k^f H_k^T (H_k P_k^f H_k^T + R_k)^{-1}. \tag{3.9}$$

   Now, every ensemble member is updated using,

   $$X_k^a = X_k^f + K_k(y_k - H_k X_k^f), \tag{3.10}$$

The analysis error covariance can be computed as,

$$\begin{aligned}
P_k^a &= (X_k^a - \bar{X}_k^a)(X_k^a - \bar{X}_k^a)^\top \\
&= (X_k^f + K_k(y_k - H_k X_k^f) \\
&- (\bar{X}_k^f + K_k(y_k - H_k \bar{X}_k^f)))(X_k^f + K_k(y_k - H_k X_k^f) - (\bar{X}_k^f + K_k(y_k - H \bar{X}_k^f)))^\top \\
&= (\mathbb{I} - K_k H_k)(X_k^f - \bar{X}_k^f)(X_k^f - \bar{X}_k^f)^\top (\mathbb{I} - K_k H_k)^\top \\
&= (\mathbb{I} - K_k H_k) P_k^f (\mathbb{I} - K_k H_k)
\end{aligned} \tag{3.11}$$

Notice that the computation of analysis error covariance does not include the factor $K_k R_k K_k$ as in Equation (3.4). This was corrected in [21] by incorporating perturbed observations in the update equation. An ensemble of observation vectors is created,

$$d_{k,p} = y_k + \gamma_{k,p},$$

where $\gamma_k \sim \mathcal{N}(0, R_k)$. $d_{k,p}$ denotes the observation vector of the p^{th} ensemble member at time t_k. We can define observation matrix $Y_k^o \in \mathcal{R}^{n \times N}$,

$$D_k = \begin{bmatrix} \vdots & & \vdots & & \vdots \\ d_{k,1} & \dots & d_{k,p} & \dots & d_{k,N} \\ \vdots & & \vdots & & \vdots \end{bmatrix}, \tag{3.12}$$

The sample covariance is computed as,

$$\tilde{R}_k = \frac{1}{m-1}(D_k - \bar{D}_k)(D_k - \bar{D}_k)^T$$

The update step, Equation (3.9) is replaced by

$$K_k = P_k^f H_k^T (H_k P_k^f H_k^T + \tilde{R}_k)^{-1}.$$

The update step, Equation (3.2) is replaced by

$$X_k^a = X_k^f + K_k(D_k - H_k X_k^f). \tag{3.13}$$

The algorithm for EnKF is shown in Figure 3.3. Note that computation of analysis error covariance is not required in the EnKF algorithm.

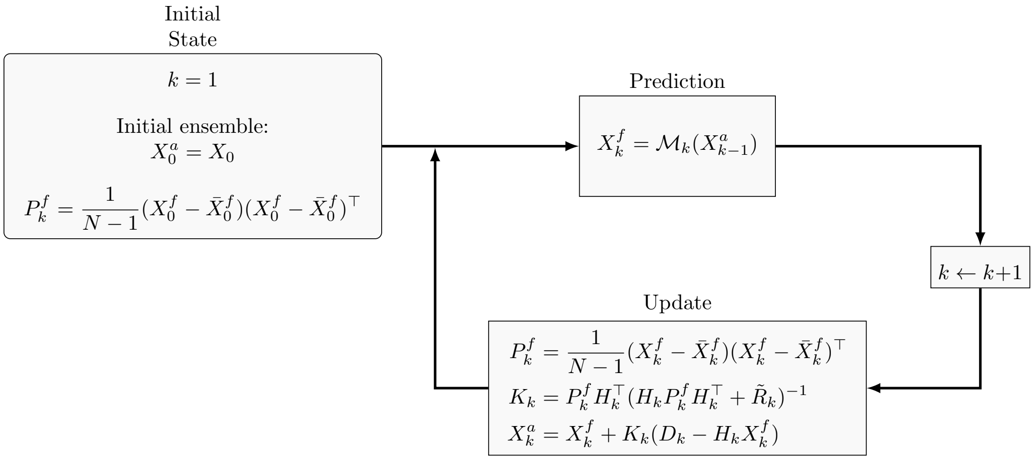


Figure 3.3: EnKF algorithm

3.3.1 Important aspects of ensemble Kalman filters and the variants

The finite ensemble size often leads to error covariance underestimation. To account for the errors due to finite ensemble size and the model errors, two common techniques are used,

3.3.1.1 Covariance inflation

The forecast error covariance is generally inflated by a factor ρ called the covariance inflation factor ($0 \leq \rho \leq 1$) [10]. Thus, the forward error covariance is modified as, $P_k^f \leftarrow (1+\rho)P_k^f$,

The size of the inflation factor generally depends on various factors such as the numerical model and is heuristic in nature.

3.3.1.2 Localization

Small ensemble size also leads to development of long-range spurious correlations in the forecast error covariance matrix[44]. As mentioned earlier, two main types of localizations are covariance localization and the domain localization. The motivation of using a localized version of filtering technique arises from the fact that observations do not affect the regions that are physically located at an unreasonable distance.

Covariance localization is performed by applying Schur product $(\circ)$ to the forecast error covariance matrix, $P_k^f \leftarrow \xi \circ P_k^f$, where, for matrices A and B, $(A \circ B)_{ij} = A_{ij}B_{ij}$. ξ is the correlation function that is non-zero in local region and zero elsewhere, as defined in [38].

Domain localization is performed by updating every point in the domain by assimilating measurement values within a radius called the localization radius. The advantage of this method is that the individual subdomains can be updated independently and hence in parallel.

3.4 Square root filters

The implementation of perturbed observations in Equation (3.13) introduces additional noise into the system and the matrix $H_k P_k^f H_k^T + \tilde{R}_k$ is rank-deficient, making inversion computationally expensive. The square root filters are employed to avoid the perturbation of observations, while ensuring that the statistics of the analysis error covariance remains consistent.

The square root filters in general, update the ensemble perturbations explicitly. Over the years, many variants of ensemble based square root filters have been developed. They differ mainly in the treatment of the forward error covariance matrices. The forecast stage for all the methods is the same as given in Equation (3.7). The difference arises in the update stage. The update stage can be viewed as the sum of two components - 1) Obtaining the mean analysis state vector 2) Updating the perturbed analysis state vector,

$$X_k^a = \bar{X}_k^a \mathbf{1_d} + \tilde{X}_k^a,$$

where $\tilde{X}_k^a \in \mathcal{R}^{n \times N}$ is the updated perturbation matrix.

Assuming we have a consistent Kalman gain matrix, the mean of analysis state vector can be extended from Equation (3.2) as,

$$\bar{X}_k^a = \bar{X}_k^f + K_k(y_k - H_k \bar{X}_k^f).$$

Next, we derive the update step of the perturbation matrix X'. We seek the analysis covariance of the form,

$$P_k^a = (\mathbb{I} - K_k H_k) P_k^f.$$

Substituting the Kalman gain matrix,

$$\begin{aligned}
P_k^a &= (\mathbb{I} - P_k^f H_k^T (H_k P_k^f H_k^T + R_k)^{-1} H_k) P_k^f \\
&= (\mathbb{I} - \tilde{X}_k^f \tilde{X}_k^{f^T} H_k^T (H_k \tilde{X}_k^f \tilde{X}_k^{f^T} H_k^T + R_k)^{-1} H_k) \tilde{X}_k^f \tilde{X}_k^{f^T} \\
&= \tilde{X}_k^f \tilde{X}_k^{f^T} - \tilde{X}_k^f Y_k^{f^T} (\tilde{Y}_k^f Y_k^{f^T} + R_k)^{-1} \tilde{Y}_k^f X_k^{f^T}, \\
&\text{where } \tilde{Y}_k^f = H_k \tilde{X}_k^f, \qquad (3.14) \\
&= \tilde{X}_k^f \tilde{X}_k^{f^T} - \tilde{X}_k^f \tilde{Y}_k^{f^T} D_k^{-1} \tilde{Y}_k^f X_k^{f^T}, \\
&\text{where } D_k^f = \tilde{Y}_k^f \tilde{Y}_k^{f^T} + R_k, \\
&= \tilde{X}_k^f (\mathbb{I} - \tilde{Y}_k^{f^T} D_k^{-1} \tilde{Y}_k^f) \tilde{X}_k^f.
\end{aligned}$$

Define a matrix $T_k \in \mathcal{R}^{N \times N}$ such that,

$$T_k T_k^T = \mathbb{I} - \tilde{Y}_k^{f^T} D_k^{-1} \tilde{Y}_k^f,$$

note that the T_k above is not unique as for any orthogonal matrix $U_k \in \mathcal{R}^{N \times N}$,

$$T_k U_k U_k^T T_k^T = \mathbb{I} - Y_k^{f^T} D_k^{-1} Y_k^f$$

and using $P_k^a = \tilde{X}_k^a \tilde{X}_k^{a^T}$,

$$\tilde{X}_k^a \tilde{X}_k^{a^T} = \tilde{X}_k^f T_k T_k^T \tilde{X}_k^f,$$

giving the update equation,

$$\tilde{X}_k^a = \tilde{X}_k^f T_k U_k,$$

3.4.1 Ensemble Transform Kalman filter - ETKF

As derived in [18],

$$T_k = C_k (\Gamma + \mathcal{I})^{-\frac{1}{2}},$$

where $C_k \in \mathcal{R}^{N\times N}$ is the orthonormal eigenvectors of $\tilde{X}_k^{f^T}\tilde{H}_k^T\tilde{H}_k\tilde{X}_k^f$ and $\Gamma \in \mathcal{R}^{N\times N}$ is the eigenvalue matrix for $\tilde{H}_k$ defined as,

$$\tilde{H}_k = R_k^{-\frac{1}{2}}H_k.$$

3.4.2 Ensemble subspace Transform Kalman filter - ESTKF

Unlike EnKF, filters such as SEIK [51] and ETKF [18] are implemented as ensemble square-root-filters that split the analysis step into a state update step and a resampling step. The state update step computes the analysis state and its error covariance matrix and the resampling step generates the ensemble members that exactly represent the analysis state and its error covariance matrix. In the current study, we have implemented the Error Subspace Transform Kalman filter (ESTKF) [82], a variant of the SEIK filter that performs ensemble transformation in the error subspace. The ESTKF combines the advantages of SEIK filter and ETKF [82]. ESTKF computes ensemble transformations that are identical to ETKF, but at a lower computational cost. Further, unlike the SEIK filter, the ensemble transformation in ESTKF is independent of the order of ensemble members in the ensemble matrix. This was demonstrated in the studies [1] where the applications are close enough to the current one. This section outlines the ESTKF algorithm that is described in detail in ([82]).

In ESTKF, the forecast error covariance matrix at time t_k is represented as,

$$P_k^f = L_k G L_k^T,$$

---

[1] `http://pdaf.awi.de/trac/wiki/WhichFiltertouse`

where,

$$L_k = X_k^f \Omega,$$

$$G = \frac{1}{N-1}\mathbb{I}_{(N-1)\times(N-1)},$$

$$\Omega_{i,j} = \begin{cases} 1 - \frac{1}{N}\frac{1}{\frac{1}{\sqrt{N}}+1} & ,i = j, i < N \\ -\frac{1}{N}\frac{1}{\frac{1}{\sqrt{N}}+1} & ,i \neq j, i < N \\ -\frac{1}{\sqrt{N}} & ,i = N \end{cases}.$$

$\Omega_{i,j}$ is the projection matrix that has a full rank, zero column sums with the columns being orthonormal. $\mathbb{I}$ is the Identity matrix. The state vector $X_k^f \in \mathbb{R}^{n\times N}$, $P_k^f \in \mathbb{R}^{n\times n}$, $L_k \in \mathbb{R}^{n\times(N-1)}$, $G \in \mathbb{R}^{(N-1)\times(N-1)}$ and $\Omega \in \mathbb{R}^{N\times(N-1)}$. The ESTKF algorithm can be outlined as follows:

1. Forecasting from t_{k-1} to t_k: Starting from a given analysis state matrix X_{k-1}^a, integrate each ensemble member $x_{k-1,p}^a$, with the dynamical system $\mathcal{M}_k$ as,

   $$x_{k,p}^f = \mathcal{M}_k(x_{k-1,p}^a).$$

2. Updating at time t_k:

   Compute the analysis state estimate $\bar{x}_k^a$ as,

   $$\bar{x}_k^a = \bar{x}_k^f + L_k A_k (H_k L_k)^T R_k^{-1}(y_k - H_k \bar{x}_k^f),$$

   where A_k is the transform matrix given by,

   $$A_k^{-1} = \rho(N-1)\mathbb{I} + (H_k L_k)^T R_k^{-1}(H_k L_k).$$

ρ is the inflation factor used to enhance the filter robustness and to partially account for model errors [87, 51]. More sophisticated methods to treat model errors have been discussed in [48].

3. Resampling:

   This step is the second order exact sampling that computes the ensemble states.

   $$X_k^a = \bar{x}_k^a \mathbf{1}_d + \sqrt{N-1} X_k^f \Omega C_k \Omega^T.$$

   where C_k is the square root of A_k.

Chapter 4

Numerical model

The Navier Stokes equations describe the physics of flow of fluids. For large scale geophysical flows (for instance tsunamis), a reasonable assumption is that the wavelength of the flow is much greater than the depth of the ocean. This assumption produces the 2D depth-averaged shallow water equations as outlined in Section 4.1. The necessary components to solve these partial differential equations include initial condition of the fluid, the topography of seabed, the boundary conditions etc.

The initial condition of the fluid is an important input. For instance, in storm surge applications, the simulations can be started using *"cold start"* or *"hot start"*. The former means that the simulation is started with the fluid at rest and the latter means that the simulation is started from a *known* state of the system. The initial condition for storm surge applications seems to be straight-forward. However, for tsunami applications, these methods cannot be used as the tsunamis are caused due to the earthquake near the fault regions at the continental shelves. The change in the topography of the seabed displaces the water, resulting to an initial water surface elevation. A typical method is to use some parametric model that can calculate the displacement of the sea bed at the start. Section 4.2 describes one such parametric model that maps the fault parameters to the topography changes.

The physical models are translated to the computer using numerical procedures such as finite element machinery. Briefly, this means that the PDEs are reformulated to their weak form and solved on a discretized space. The physical domain is discretized into cells and the PDEs are solved weakly at every cell. Finite volume method is one such technique, wherein the solution at a given cell is constant valued. Section 4.3 describes the numerical package GeoClaw, that uses finite volume method to solve hyperbolic equations described below.

4.1 Shallow water equations

Given that the physical coordinate space is aligned in x and y directions, the shallow water equations are given as,

$$h_t + (hu)_x + (hv)_y = 0, \tag{4.1}$$

$$(hu)_t + (hu^2 + 1/2gh^2)_x + (huv)_y = -ghB_x - Du, \tag{4.2}$$

$$(hv)_t + (huv)_x + (hv^2 + 1/2gh^2)_y = -ghB_y - Dv, \tag{4.3}$$

where $h(x, y)$ is the total depth of water column, $u(x, y, t)$ and $v(x, y, t)$ are depth averaged velocities in the x and y directions respectively, $B(x, y, t)$ is the bathymetry, $D(h, u, v)$ is the drag term and g is the acceleration due to gravity. The drag term used is of the form,

$$D = \frac{gM^2\sqrt{u^2 + v^2}}{h^{\frac{5}{3}}},$$

where M is the Manning's coefficient. The Manning's coefficient depends on the substrate surface and is empirically computed. For the current work it is assigned a value of 0.025. Equation (4.1) represents the conservation of mass. Equation (4.2) and Equation (4.3) represents conservation on momentum in x and y directions, simultaneously. Additionally, Coriolis forcing can be added to the momentum equations. However, for tsunami applications, the contributions are negligible.

4.2 Tsunami applications

The initial condition of the system is a necessary input for solving the partial differential equations. For tsunami models, the initial water surface elevation is generated from the bathymetry changes due to the associated earthquake. The earthquake source models typically specify the slip along the fault or on subfaults making up a single place. One such parametric model given by [84] relates the topography change Δd, to various fault parameters,

$$\Delta d \sim f(strike,\ length,\ width,\ depth,\ slip,\ rake, dip,\ longitude,\ latitude),$$

where the parameters are defined as follows,

1. *length* and *width*: Length and width of the fault place (typically in meters or kilometers).
2. *latitude* and *longitude*: Latitude and longitude of some point on the fault plane, typically either the centroid or the center of the top.
3. *depth*: Depth of the specified point below the sea floor
4. *strike ($0°$ to $360°$)*: The orientation of the top edge, measured in degress clockwise from North.

5. *dip (*$0°$ *to* $90°$*)*: Angle at which the plane dips downward from the top edge.

6. *rake (*$-180°$ *to* $180°$*)*: Angle in the fault plane in which the slip occurs, measured in counterclockwise from the strike direction.

7. *slip (*> 0*)*: The distance (typically in centimeter or meter) that the hanging block moves relative to the foot block in the direction specified by rake. The hanging block is the one above the dipping fault place (or to the right if you move in strike direction).

The Okada model is an approximation to the seafloor deformation because of the underlying assumption that the seafloor is flat and the deformation occurs in a solid medium that is homogenoeus isotropic elastic material. However, this is a reasonable assumption as the fault parameters are often not known very well and contain uncertainties.

4.3 GEOCLAW

Equation (4.1), Equation (4.2) and Equation (4.3) are solved using GEOCLAW ([16]), a part of the general Clawpack (Conservation Laws) package [72] designed to solve wave propagation problems. Clawpack uses high-resolution shock-capturing finite volume methods on logically Cartesian grids to solve non-linear hyperbolic systems of partial differential equations [63]. GEOCLAW is written in Fortran , with a user interface in Python. The Clawpack package uses adaptive mesh refinement. The adaptive mesh refinement aids in obtaining high-resolution solutions on concentrated regions of the domain. For instance, for tsunami applications, the regions of interest might be the traveling wave front, the coastline areas or the islands in the path of the wave front. The adaptive mesh refinement algorithm dynamically refines the mesh (based on certain criteria) to provide solutions of kilometers scale in deep ocean and meter scale at the coastlines.

4.3.1 Adaptive mesh refinement

The AMR algorithm used in GEOCLAW is described in detail in [63, 64, 39, 16]. This section will briefly describe the algorithm.

AMR is implemented using a patch-based mesh refinement approach. As shown in Figure 4.1, multiple levels of structured meshes are used to reach a fine resolution [17]. For notation, a $Level\ k$ refinement is called a k^{th} level patch. The coarsest mesh is the $Level\ 1$ patch that always exists by default. At a given level, there may be multiple finer level regions called *grids*. Thus in Figure 4.1, Level 2 patch has 1 grid.

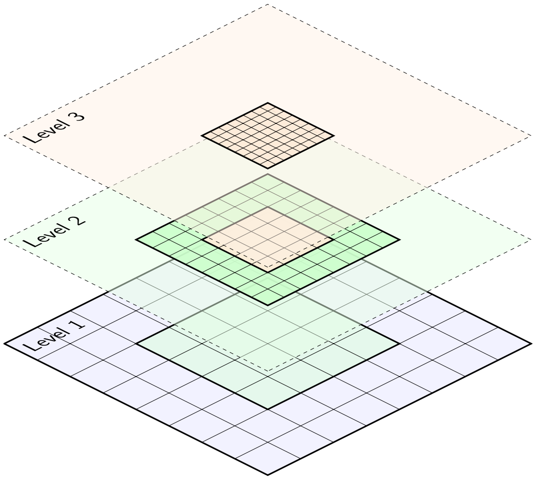


Figure 4.1: Adaptive mesh structure in GeoClaw: A single Level 1 grid covers the entire domain. Some rectangular portions of this grid are covered by Level 2 grids refined by some refinement factor in each direction. Regions of each Level 2 grid may be covered by Level 3 grids, that are further refined, and so on.

Refinement to higher levels is done by first flagging the cells at the coarsest level. The cells are flagged based on various criteria such as topography, wave speed, initial water height, water surface elevation above a specified level above sea level among others. The flagged cells are then clustered into rectangular regions where further refinement takes place. The clustering algorithm is based on constructing the most *efficient* rectangular regions. The criteria is that the rectangular grids do not contain many unflagged cells while the total number of mutually exclusive grids at a given level are minimum. Once the finer level grids are constructed, the numerical solution is obtained on the finer mesh. For stability of the explicit finite volume scheme, the time step on the refined patches is also refined to satisfy the *CFL* criteria. Additionally, for tsunami modeling, GeoClaw has the capability to use anisotropic refinement in time as well. The gridding operation from coarse mesh to a finer mesh is recursively performed till the highest possible refinement level.

After every few time steps, the entire operation (also called as *regridding*) is performed. The regridding operation allows to follow the features of interest as the wavefronts travel. Further, the regridding operation coarsens the mesh in regions where refinement is unnecessary.

Chapter 5

Coupling ensemble based filters and GEOCLAW

A data assimilation system is a combination of the forecasting model coupled to the assimilation technique/filter. Chapter 3 describes the ensemble based data assimilation techniques and Chapter 4 describes the numerical model and its implementation using GEOCLAW. This section is the crux of the research and highlights various contributions. The objective of the research is to develop a data assimilation system coupled with a model that has adaptive mesh refinement feature.

Here the coupling means to develop an integrated system such that the numerical model performs its task of forecasting the ensemble members till the update step. The solution of the various ensemble members are passed to the data assimilation filtering library or a program that essentially updates the ensemble members and provide the assimilated members as output. This sequential approach of forecasting and update takes place by communicating the quantity of interest (in our case water surface elevation) between the numerical model and the assimilation program or the library.

---

[0] Pushkar Kumar Jain, Kyle Mandli, Ibrahim Hoteit, Omar Knio, and Clint Dawson. Dynamically adaptive data-driven simulation of extreme hydrological flows. Ocean Modelling, 122:85 – 103, 2018. The specific contribution includes development of the coupling mechanism known as mesh union strategy and the various artifacts of the data assimilation system it entails.

There are two ways to couple a filtering technique to a numerical model. They are the 1. Offline method and 2. Online method. The offline method treats the numerical model and the assimilation technique as individual black boxes. At every update step, the forecasted solution of various ensemble members are reformatted into a suitable format that is compatible to the assimilation program or the library. After the update, the solution of various ensemble members is reformatted back to the format that is compatible to the numerical model, thereby, making it ready for the next forecast stage. The advantage of this method is that little knowledge of the individual black boxes is required and the assimilation system can be set up without any complications that would otherwise be encountered in modifying the source code. The disadvantage of this method is the additional inevitable overhead that is required to read and write the data back and forth between the individual black boxes. For real-time practical applications, where time is critical, this can be a serious bottleneck.

The online method constructs a single computer code that integrates the functionalities (API) of the filtering technique directly into the existing numerical model. The advantage of this method is that the data communication between the two systems is seamless. The disadvantage includes deeper understanding of the internals of the numerical model and the filtering technique and possibility to easy code modification. Note that, typically for large scale geophysical models, the codes use parallelism such as MPI or OpenMP. Hence, implementation of online method can be challenging. For the current work, the online mode has been chosen.

This chapter describes the challenges faced in the development of an online coupled system of GEOCLAW with the ensemble assimilation techniques and proposes a solution. Section 5.1 describes the algorithm addressed a "supermesh" methodology, for handling the ensemble members with different AMR structures at the assimilation step. Independent of the numerical model, the components of ensemble based filters are the generation of initial ensemble members, state vector that will be assimilated (5.2) and the observation operator that maps the measurements from state space to observation space (5.3). During the course of the research, it was found that special attention was required to address the generation of ensemble members with respect to tsunami models. Hence, this is separately dealt with in Section 5.4. Finally, Section 5.5 describes the parallelism in the coupled system.

5.1 Mesh union strategy

The primary difficulty in using AMR in conjunction with ensemble filtering schemes is that the regridding operation during the forecast allows every ensemble member to adapt its grid in its own way. Hence, at the end of the forecast stage, one ensemble member may obtain a grid structure that is different from the other ensemble members. For instance, it is possible that at the end of the forecast stage, at least one ensemble member may have a $Level$ 3 grid in some regions of the domain, while other ensemble members may have just $Level$ 2 grid in the same regions. Figure 5.1 illustrates a hypothetical scenario of two ensemble members at an analysis stage, with different refined patches at level 2 and level 3. The grids in Figure 5.1 also list the number of cells at each level for the ensemble members. The first ensemble member has two disjoint patches at level 2, while the second ensemble member has a level 2 patch in a different part of the domain. Further, there exists different level 3 patches. Thus, at finer levels, the ensemble members have varying number of cells and thereby end up with a varying state vector length. This means that statistical quantities such as covariance of the state, essential for the update step, are no longer straightforward to compute.

This motivates us to develop a methodology that can enforce all the ensemble members to produce the same mesh structure at all levels at the analysis stage. In a nutshell, the method is to (1) construct a union of meshes of all the ensemble members, also called a *supermesh*, (2) communicate the information of every ensemble member to the supermesh, (3) perform analysis, (4) communicate the updated information back from the supermesh to the respective ensemble member, and (5) continue the forecast.

The following four remarks highlight the intricacies of implementing the above methodology. It is found that even though the mesh union algorithm seems straight-forward, certain characteristics of the adaptive mesh algorithm makes this a complicated procedure during implementation.

Remark i - As mentioned, after every few time steps, GEOCLAW performs a regridding operation to adapt the mesh to the features of interest. Construction of a supermesh is done by modifying the default regridding operation to a constrained regridding operation. The mesh structure of each ensemble member is communicated to every other ensemble member. Thus, at any given point in the domain, we know the maximum refinement level that exists for at least one ensemble member. Formally, let $\phi^{(i)}_{(x,y)}$ and $\delta^{(i)}_{(x,y)}$ be the maximum refinement level at a point (x, y) for the i^{th} ensemble member, before and after the analysis step respectively. That is,

$$\delta_{(x,y)}^{(i)} \triangleq \max_{i} \left(\phi_{(x,y)}^{(i)}\right).$$

After the analysis step, the mesh structure for all the ensemble members is the same. Hence, $\delta_{(x,y)}^{(i)} \equiv \delta_{(x,y)}$. This way we ensure that during the assimilation, all the members have the same mesh structure at the start of the assimilation step.

Remark ii - AMR algorithm in GEOCLAW constructs a finer mesh to Level $k+1$ by first flagging cells in Level k. At Level k, the flagged cells are clustered based on a clustering parameter that defines the cut-off percentage of flags required to bound the flags within a rectangular region and construct a patch [15]. This implies that the union of two ensemble members may not lead to an "exact union“. Nevertheless, since all the ensemble members have the same flags at all levels, the deterministic nature of the clustering algorithm guarantees the same finer patches for the finer levels. When this procedure is applied recursively from coarsest mesh to finer mesh, all the ensemble members end up with the same mesh structure.

Remark iii - In the context of an MPI environment, where a single processor is forecasting a single ensemble member, the mesh structure of every ensemble is communicated to the ensemble member through an "MPI_GatherAll“ operation. Development of a supermesh implies that all the ensemble members may obtain an over-refined mesh structure in the domain. However, this is only at the analysis step. A regrid operation in the forecast step (after assimilation) removes the unnecessary refinement and the ensemble members undergo forecasting in the usual manner thereafter.

Remark iv - In practice, it is observed that the mesh structures of the ensemble members do not vary in high degree from each other. An ideal way of execution of the above step would be to 1) gather the mesh structure of all ensemble members to a single processor, 2) construct a maximum disjoint set of the rectangular regions from all the ensemble members, and 3) broadcast the new mesh structure to all the processors. Thus, the communication cost can be reduced at the analysis step.

Implementation of the above strategy of the union of meshes leads to the final mesh shown in Figure 5.1, on which assimilation will be performed. It should be noted that a perfect union may not occur due to the nature of the AMR algorithm used, as mentioned in **Remark ii**. As an illustration, Figure 5.2 demonstrates the mesh union operation for a two ensemble system.

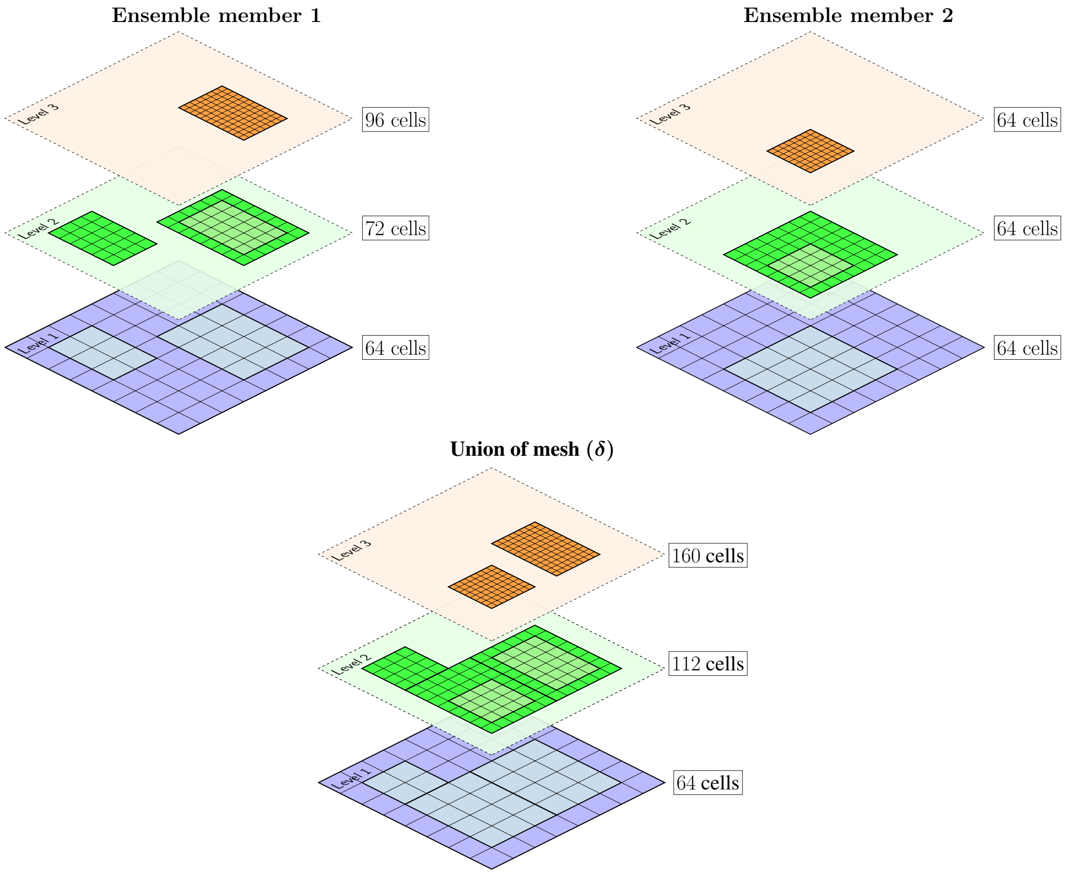


Figure 5.1: Hypothetical mesh union operation on 3 level AMR mesh. Note that in the super-mesh, the Level 3 mesh has disjoint patches. AMR algorithm partitions the Level 2 accordingly as well.

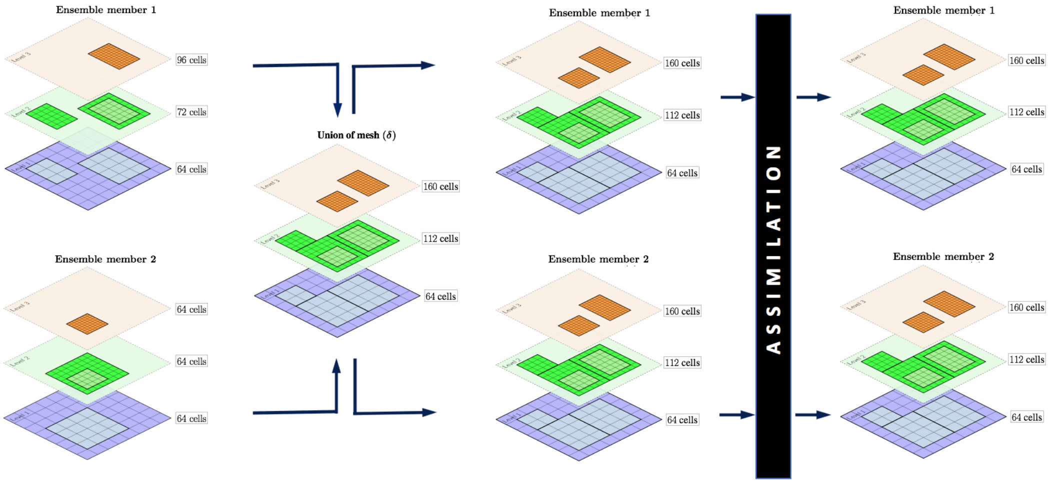


Figure 5.2: The mesh union algorithm for two ensemble member system

Here we illustrate the generation of the supermesh grid via a GeoClaw run of a model problem that simulates the water flow in a parabolic bowl. The initial condition of the water surface elevation is a Gaussian hump shown in Figure 5.3. We run the simulation for 3.2 seconds. The maximum AMR level that is possible during the simulation is set to 2. We generate two ensemble members with different initial amplitudes. Figure 5.4 shows the snapshots of water surface elevation at instances when the supermesh is formed. The left and the middle column represent the two ensemble members, and the right column is the case if the supermesh was formed. In the figure, the rectangular patches without any cells (clean patches) are the level 2 patches. It is seen that from 2.56 seconds, as the water front is moving radially outward, different initial amplitudes causes the ensemble members to adapt with different meshes, forming a supermesh accordingly. As mentioned in Remark ii, the supermesh grid structure is not a perfect union of grids of the ensemble members.

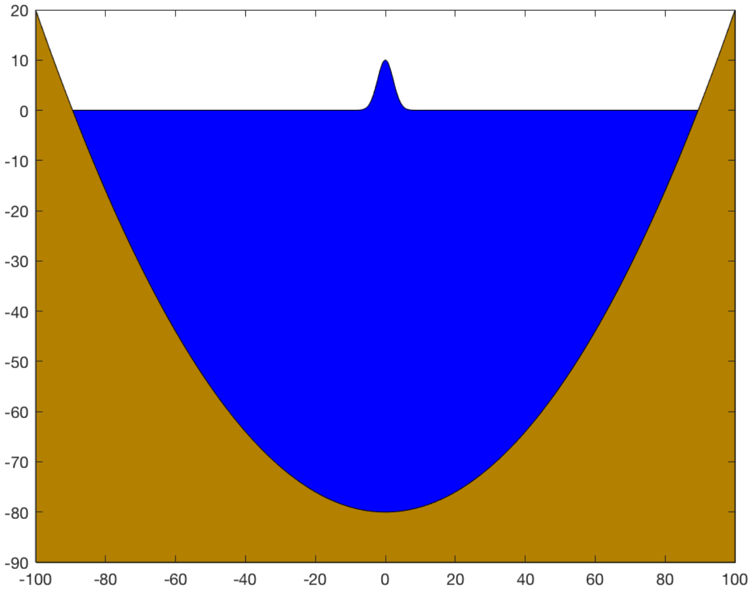


Figure 5.3: Cross-sectional view of the initial water surface elevation (blue) and the parabolic bowl topography (brown) for one of the ensemble members in the experiment to illustrate generation of supermesh.

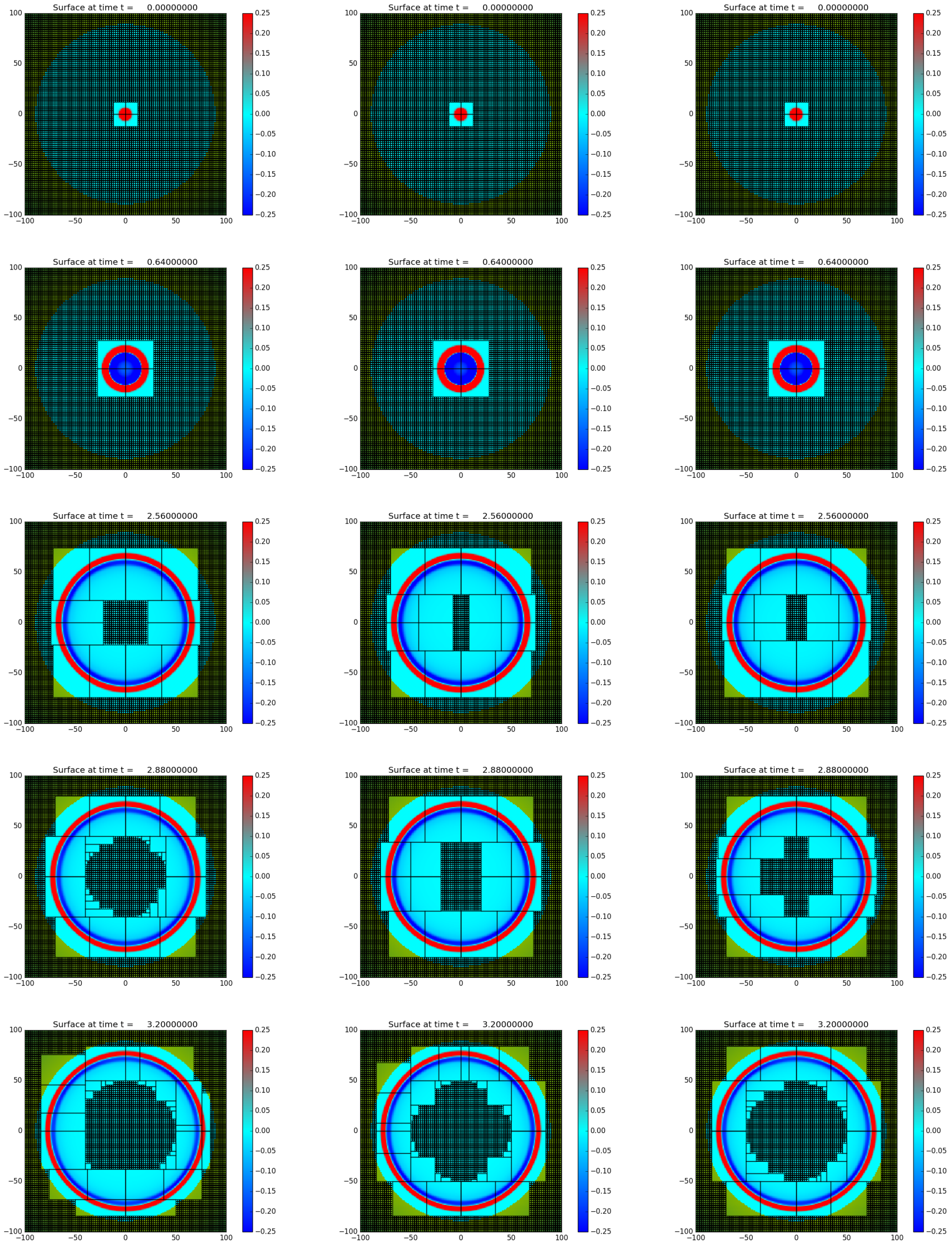


Figure 5.4: Radial bowl simulation showing generation of supermesh at the assimilation stage. Left - Ensemble member 1; Middle - Ensemble member 2; Right - Ensemble union generated

5.2 Construction of state vector

For the current study, the state vector is constructed in a way that all the cells at all the levels of the supermesh are combined into one single array. We define a patch as the collection of grids at a given level. If P is the maximum refinement level in δ, then we define the state to be assimilated (ψ), water surface elevation as

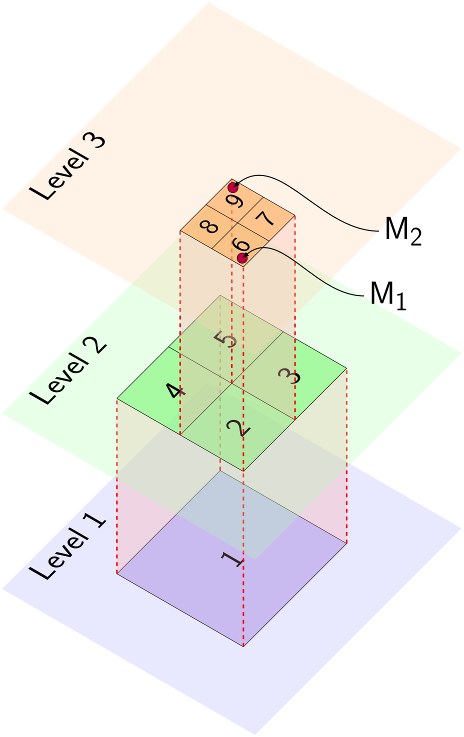
$$\psi \triangleq \begin{bmatrix} level\ 1\ patch \\ \vdots \\ level\ k\ patch \\ \vdots \\ level\ P\ patch, \end{bmatrix},$$

where we have augmented the state vector at various levels into a new state vector. For instance, the state vector constructed from the hypothetical union of mesh in Figure 5.1 would be of size $64 + 112 + 160 = 336$. This means that if our state vector has a cell value at a finer level, then it also contains the underlying coarser cell value.

5.3 Construction of observation operator

In a multi-level AMR structure, a measurement may come from a location where more than one level of refinement exists. A measurement recorded at a physical location that belongs to the finest cell will also belong to the underlying coarse cells. Given that the state vector combines the variables from all the levels at a given physical location, the observation operator is constructed in such a way that the same observation realization is used in all the levels at that physical location.

For instance, consider the mesh structure in Figure 5.5 that represents a single coarse cell denoted by v_1 at Level 1. The coarse cell v_1 is refined to finer cells denoted by v_2, v_3, v_4, v_5 at Level 2. The cell v_2 is further refined to cells v_6, v_7, v_8, v_9 at Level 3. The state vector for this simple structure is $\psi = [v_1, v_2, \ldots, v_8, v_9]^T$. Depending on the location of the incoming measurement, multiple scenarios occur. **Scenario 1** - If there exists just a single incoming observation M_1 at a point belonging in cell v_6, then clearly, M_1 belongs to v_2 and v_1 as well. Hence the measurement recorded for v_1, v_2 and v_6 is M_1. **Scenario 2** - If there exists two incoming observations M_1 and M_2 at the locations belonging to cells v_6 and v_9, respectively, then both the measurement values belong to v_2 and v_1 as well. Hence, the measurement values recorded at v_6 is M_1, at v_9 is M_2 and at v_1 and v_2 is taken as the average of M_1 and M_2. This idea can be extended to other scenarios as well.



;

Figure 5.5: A simple 3 level mesh structure representing a state vector size of 9 elements along with 2 incoming observations M_1 and M_2.

5.4 Data assimilation for tsunami applications

5.4.1 Motivation

Initial ensemble generation can prove to be a challenging part for data assimilation coupled to tsunami models. Generating a representative initial ensemble to start the assimilation systems is important to obtain good performance with ensemble based filters. Several approaches have been proposed [101, 49, 86]. For instance, [86] discusses the idea of generating the initial ensemble for large scale ocean and atmosphere models via second-order exact sampling from a model trajectory that is obtained from years of spin-up. However, unlike the ocean and atmospheric models, where the initial ensemble of states can be produced via model spin-ups, tsunamis occur relatively instantly due to the rapid bathymetry changes. We address this inability to generate ensembles via model spin-ups by an algorithm that can effectively map the uncertainty in the bathymetry changes at the event of tsunami to the initial ensemble.

Section 4.2 discusses the Okada model that maps fault parameters to the topography change. In reality, the fault parameters are seldom known perfectly. Further, at the time of such extreme events, efforts are more concentrated towards forecasting tsunami models as soon as possible and assessing the possible damage that may be caused in the next few hours.

The idea is to map the uncertainty in the fault parameters to an ensemble of initial states of the tsunami. The first approach was to take a sample of the fault parameters such that each perturbed state would produce an ensemble member. However, it was observed that this method yielded unstable solutions near the fault regions. The reason is that every ensemble member has a different bathymetry near the fault regions that exists at the continental shelf and the update step results in flooding of land regions that are much away from the coastal regions. This problem do not arise for earthquake locations that are deep in the ocean (away from the land regions). One possible remedy to this is having a localized update of the state near the fault regions such that only the state that is in the ocean regions are initially updated. However, this is cumbersome and may require manual intervention on case-by-case basis.

This motivates to develop a procedure that can produce the initial ensemble members from the uncertainty in the fault parameters. The idea is to use the uncertainty in the fault parameters to compute the subspace of the ensemble members. Then ensemble members can then be sampled from this subspace.

5.4.2 Procedure

We assume that the fault parameters have a multivariate Gaussian distribution with a specified mean and a covariance. We also assume that the fault parameters are not correlated to each other and hence the covariance matrix is diagonal with each entry representing the variance of a fault parameter. Under these assumptions, first perturb the fault parameters using a relatively large sample N_s. The subspace of the perturbations generated, represents the possible bathymetry deformations. Thus, a forward integration of the model up to a small time (called checkpoint time) with each of the samples as the initial condition, generates the subspace of initial tsunamis. Similar to the idea proposed by [86, 49], we adopt second-order exact sampling from various model states. But unlike the temporal snapshots of the states, we use the snapshots of the various possible tsunamis that we have generated, to develop a perturbation matrix of a desired size. These perturbations can then be later added to the mean water surface elevation at the checkpoint time and in turn generate the initial ensemble.

Using the nomenclature defined in Chapter 3, we describe the procedure to generate the perturbation matrix. Obtain the matrix $A = X - \bar{X}\mathbf{1_d}$. A represents the variability among the states at the checkpoint time. Compute the covariance matrix as $A^{'} = (N-1)^{-1}AA^T$. An eigen decomposition of $A^{'}$ gives the eigenvalues and the associated eigenvectors or the empirical orthogonal functions (EOF). The eigenvectors associated with the associated large eigenvalues represent the significant directions that have the maximum variability. Hence, perturbations in these directions would yield the most effective ensembles. The singular values and the singular vectors of $A^{'}$ can be used to compute a perturbation matrix of the desired ensemble size, that holds the perturbations along the columns. For this, we obtain the product of every singular value with the corresponding singular vector [51]. We then multiply the matrix with a random matrix that preserves the mean and the covariances. This random matrix can be generated using Householder reflections. Multiply the result with $\sqrt{N-1}$. The resulting matrix holds the ensemble perturbations[86].

5.4.3 Implementation

Large scale ocean model softwares use checkpoint and restart capabilities, wherein, a forecast can be resumed from a checkpoint time. We rely on this capability to implement the initial ensemble generation algorithm. Let us denote t_o, t_{cp} and t_{end} as the initial time, the checkpoint time and the simulation final time respectively.

1. Perturbation generation (Checkpoint run) - At time $t_o = 0$, draw a relatively large random sample (say $N_s = 100$) of the fault parameters of the Okada model, under the assumption that the parameters are uncorrelated to each other and are normally distributed with a user specified mean and a diagonal

covariance. Integrate using each of the fault up to a reasonably small time, called the checkpoint time t_{cp} to generate the subspace of possible tsunamis.

2. Use the 100 model states (N_s) and a user desired ensemble size (n_s) to generate the perturbation matrix. Add the perturbations to the mean water surface elevation at time t_{cp}.
3. Assimilation run - With the assimilation facility activated, resume the code at the checkpoint time t_{cp} till the final time t_{end}.

Remark i - For a given problem, the eigen decomposition has to be performed only once at the initial time $t_o = 0$. The eigenvectors and the values are small in size and can be stored on disk. This can be a part of the preprocessing stage.

Remark ii - A small checkpoint time t_{cp} implies that forecasting N_s (here 100) ensemble members till the checkpoint time is not so computationally expensive.

5.5 Development of data assimilation framework

We have used the data assimilation software - Parallel Data Assimilation Framework (PDAF) [79], a Fortran library for various ensemble filters such as EnKF, ETKF, ESTKF, SEIK, SEEK and the localized filters such as LETKF, LSEIK, LESTKF. The library supports parallel ensemble forecasts and assimilation in MPI and OpenMP. As seen in Figure 5.6, each GEOCLAW run is a serial run representing a single ensemble member. The forecast of the ensemble members are executed in parallel under MPI parallelism. For the assimilation stage, the forecasted state vectors are gathered to the root for the update. The updated states are scattered back to the cluster for next forecast stage. For the current study, a single ensemble member is assigned to a single core during the forecasting. Hence, a total of N ensemble members is performed by N cores in a cluster, wherein, each core is assigned to one ensemble member. The assimilation is performed by a single core as well. It should be noted that the PDAF library has the capability to adjust the number of cores per ensemble member during the forecast stage and the number of cores in the update step. The simulations were carried on a local desktop with 8 cores and not in a distributed environment. A virtual MPI environment was created such that the 8 cores would mimic the user-defined number of cores. Hence, the wall clock time will not be a good estimate for analyzing execution time. In the present study, the coupling is done in such a way that the forecasts are distributed amongst the processors and the analysis is done by a single processor.

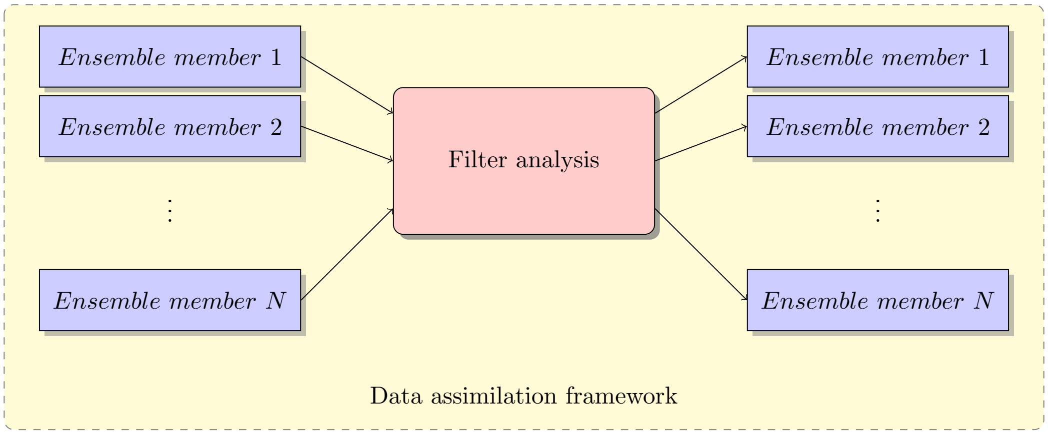


Figure 5.6: Parallelism in GeoClaw-PDAF data assimilation framework

Chapter 6

Numerical results

Chapter 5 describes the development of data assimilation system with the numerical model GEOCLAW that uses adaptive mesh refinement.

Development of verification test cases can be a challenging task for testing assimilation systems. The methodologies typically involve *twin* experiments where, one simulation represents the *truth* and the other simulation represents a *biased* run that assimilates measurement data from certain locations of the *truth*. Three test cases are discussed - *Radial bowl test case*, *Chile tsunami of February 2010* and *Tohoku tsunami*.

In real world problems, the initial condition of the system is seldom known. As discussed earlier, this is certainly a problem for tsunami applications where the exact bathymetry changes due to the associated earthquake is impossible to know. Hence, there are always uncertainties in the initial conditions. Additionally, there are other errors such as the discretization errors. The numerical solution is obtained on a discretized space and the numerical method used contains errors due to the

---

[0] Pushkar Kumar Jain, Kyle Mandli, Ibrahim Hoteit, Omar Knio, and Clint Dawson. Dynamically adaptive data-driven simulation of extreme hydrological flows. Ocean Modelling, 122:85 – 103, 2018. The contribution includes application of data assimilation system on various test cases as part of the Chile tsunami event.

approximations.

The *Radial bowl test case* addresses data assimilation reducing the discretization error and the initial condition error.

The *Chile tsunami* test case discusses the use of data assimilation towards tsunami application problems. We look into the aspects of localization radius, observation configurations, minimum ensemble size and the initial ensemble affecting state estimates. This is followed by an AMR-coupled assimilation run of the observed real event, taking into account the available gauge data. This test case highlights on calibration of the model for effective assimilation. The lessons from this are implemented in the *Tohoku tsunami* test case.

The *Tohoku* test case serves as as a validation test case. Gauge data is available at certain locations. The best practices from *Chile test case* are imbibed in this test case. The validation is performed by assimilating some gauge measurements and observing improvement in errors in other gauges.

All the three test cases are discussed further in the subsequent sections.

6.1 Radial bowl test case

The radial bowl test case simulates the flow of water in a parabolic bowl such, wherein the initial condition is a gaussian hump of water that is solely forced by gravity. For the various experiments explained in further sections, the test setup is defined in Table 6.1.

Table 6.1: Model parameters for simulating radial bowl test case

Computational domain	$x \in [-100, 100]$ and $y \in [-100, 100]$
Simulation time	$9.6\ seconds$
Time stepping	Initial time step $= 0.016\ seconds$ and desired $cfl = 0.75$
Topography	Parabolic bowl, $z(x, y) = 0.01(x^2 + y^2) - 80.0$

6.1.1 Assimilation and discretization error

We define the discretization error as the difference between the solution on a high-resolution mesh with the solution on a coarser mesh. A higher resolution of the mesh (as the grid gets finer) generates a solution that is more accurate and detailed than the same simulation on a coarser mesh. This means that if the solution on the coarser mesh is assimilated with solution from the finer mesh as observations, a successful assimilation would update the coarse cells with state values from finer mesh and thereby reduce the discretization errors.

True solution: Solution on a fine level mesh. Note that this can be substituted with the analytical solution as well. For the current run, true solution is generated on third level of refinement

Free run solution: Solution on the coarse mesh without any assimilation

Assimilated solution: Solution of a simulation on the coarse mesh with assimilation. The assimilation parameters are given in Table 6.2.

Table 6.2: Assimilation parameters for simulating radial bowl test case to test the effect of assimilation on discretization error

Number of ensembles	30
Localization radius	20°
Assimilation steps	Every 0.32 seconds
Standard deviation in ensemble generation	0.5 meters
Observation configuration	Figure 6.1
Inflation factor	0.8

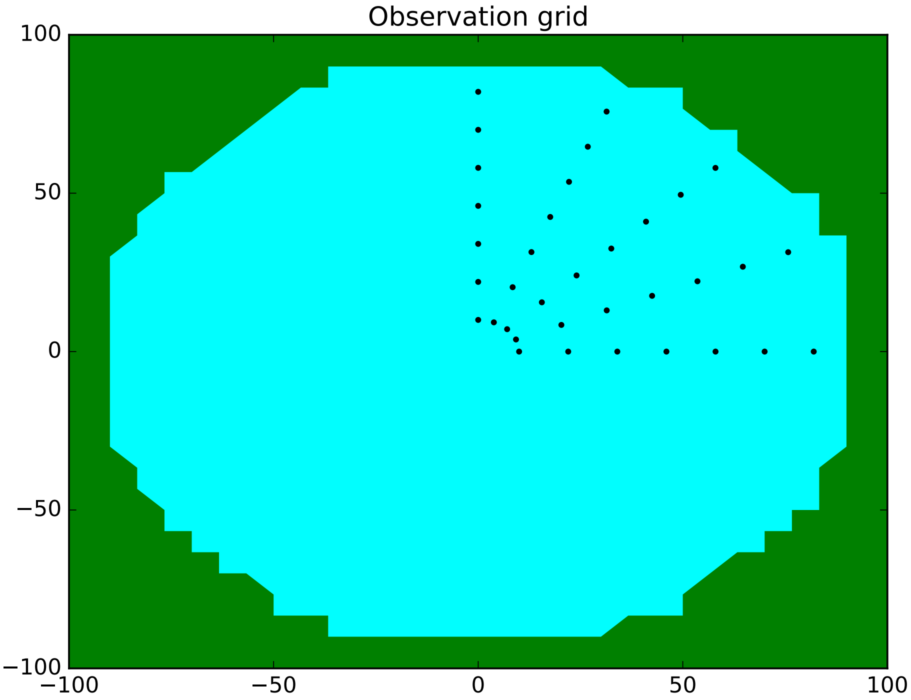
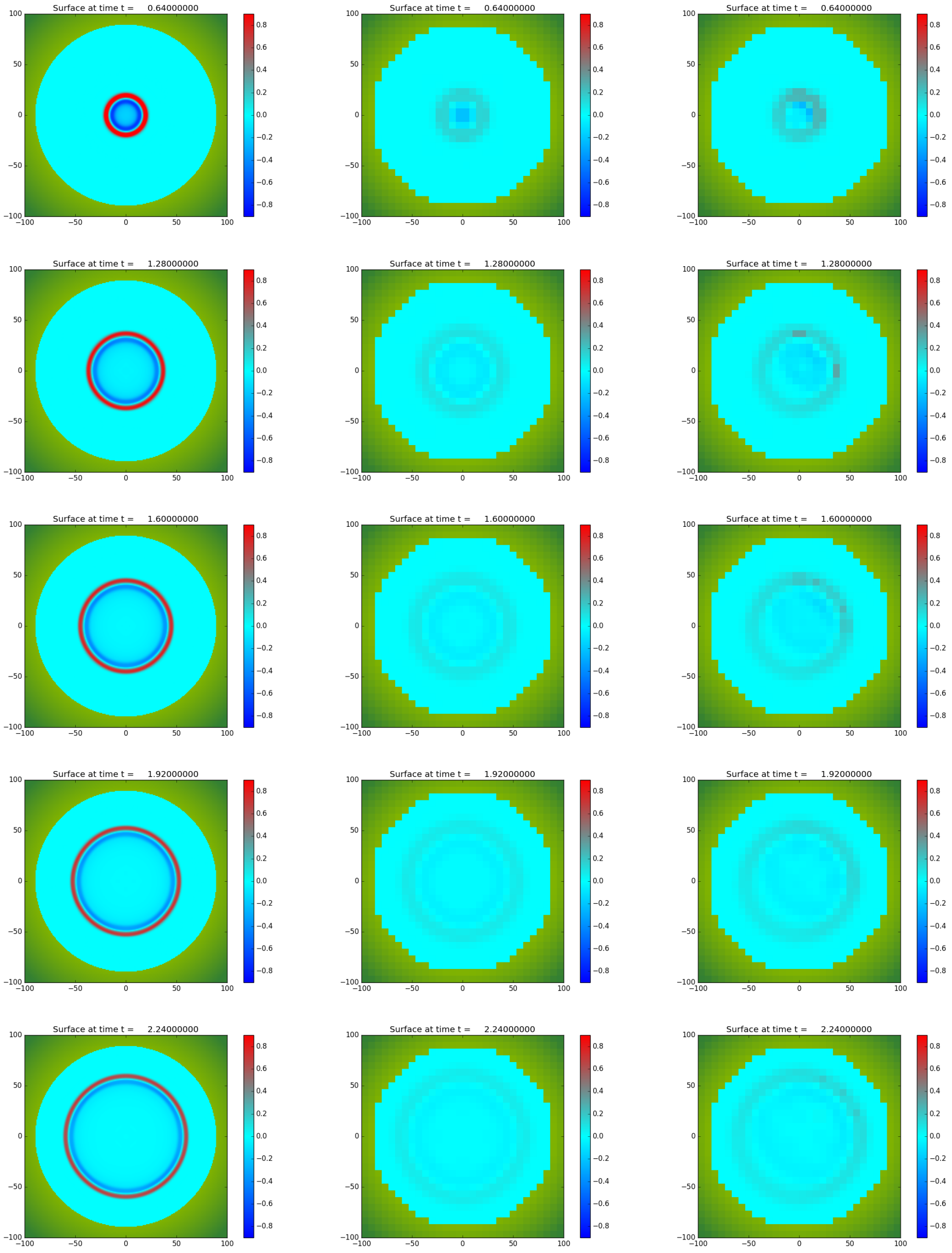


Figure 6.1: Radial bowl test case observation configuration

Error evaluation: The error metric used is the pointwise root mean square of the difference in the water surface elevation between the true solution and the solution of interest (free run or assimilated solution). As per the hypothesis, the errors of the assimilated solution must be lesser than the free run simulation.

Observation configuration: The observation configuration is shown in Figure 6.1. The observation configuration was motivated by the fact that the update is more pronounced as the leading wave approaches the boundary. Note that other observation configurations were also analyzed. Overall, the conclusion of the results remains the same.

Three simulations are run to generate the truth, free run and the assimilated solution. The input conditions for all the simulations remain the same. This is required to explicitly account for decrease in the discretization error to assimilation effect. Figure 6.2 shows the water surface elevation plots for the three different runs. The range of color bar is -0.8 to 0.8 meters. The left column is the true solution on a Level 3 mesh. The center column is the water surface elevation on the coarse mesh. The right column is the same run with assimilation using observations from the true solution projected on the coarse mesh. It is seen that the free run has a lower amplitude of the leading wave. This is corrected in the assimilation run. This is further seen in the error plots in Figure 6.3 and Figure 6.4, where the error between the true solution and the assimilation solution is lesser than the error between the true solution and the free run.



(a) Truth (b) Free run (c) Assimilated

Figure 6.2: Water surface elevation snapshots for radial bowl test case to test assimilation aiding in reduction of discretization error. Left - *Truth*; Center - Free run 2; Right - Assimilated

Effect of observation rms error: As seen from Figure 6.3, as the certainty in the observation values increases, the update values are closer to the truth and hence visible in decreased error values. Though in the current case, the true values are very certain, in reality extreme reliance on the sensor data may be harmful and produce erroneous results if the sensors record any anomalous values. A reasonable rms error value that works well for the test cases is meters.

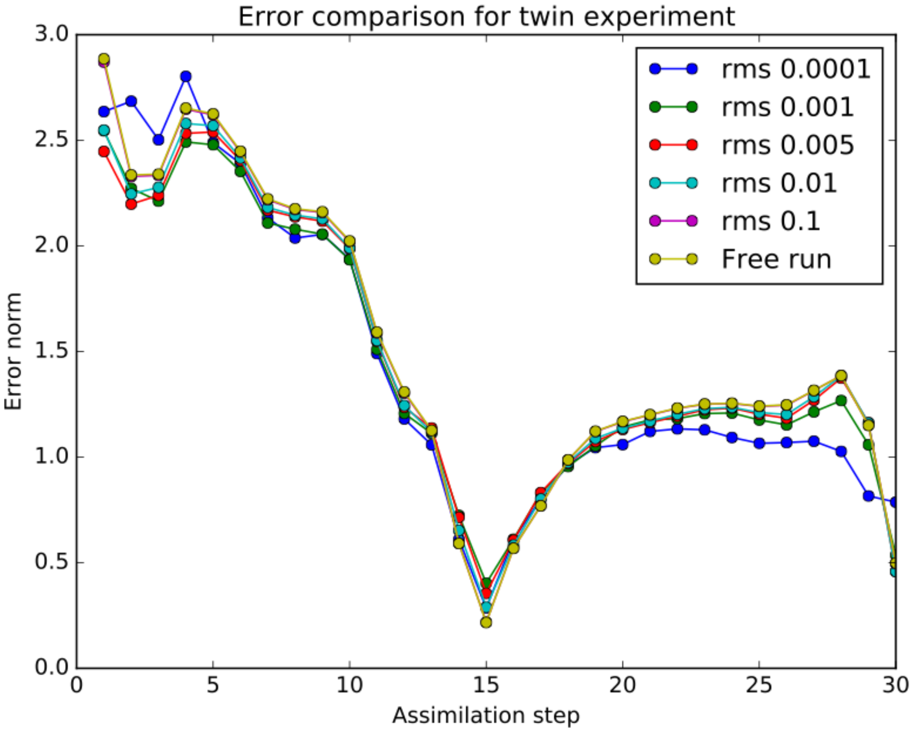


Figure 6.3: Error comparison between true solution, free run and the assimilated solution for testing effect of observation rms error on assimilation.

Effect of filter type: As seen from Figure 6.4, the filters perform the same till the 15 assimilation steps. This corresponds to the leading wave reaching the boundary. However, it is observed that the errors in the returning wave is the least for localized SEIK filter.

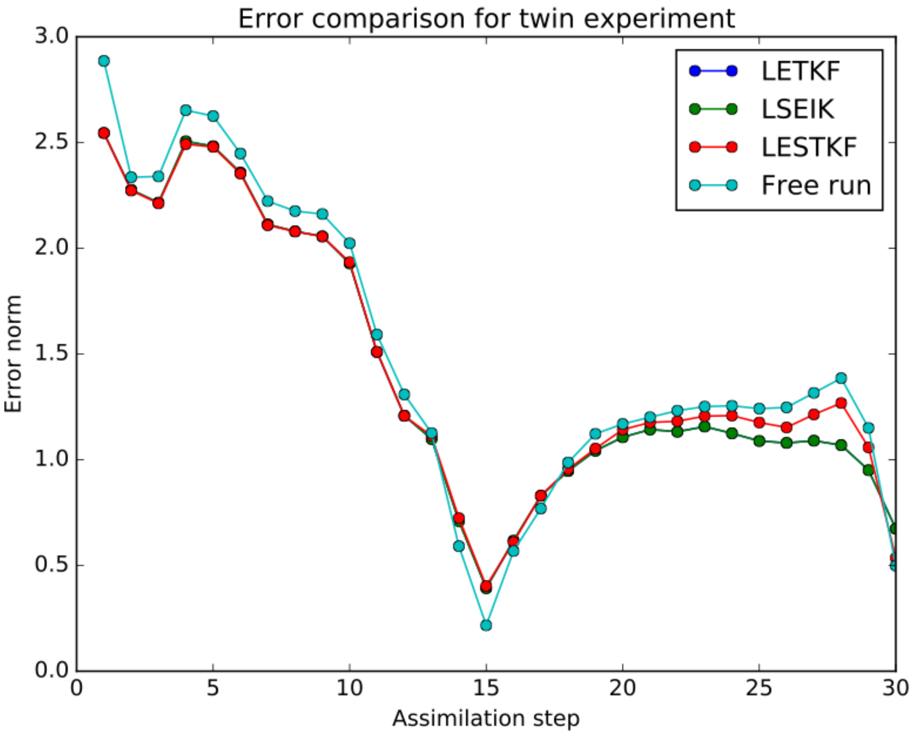
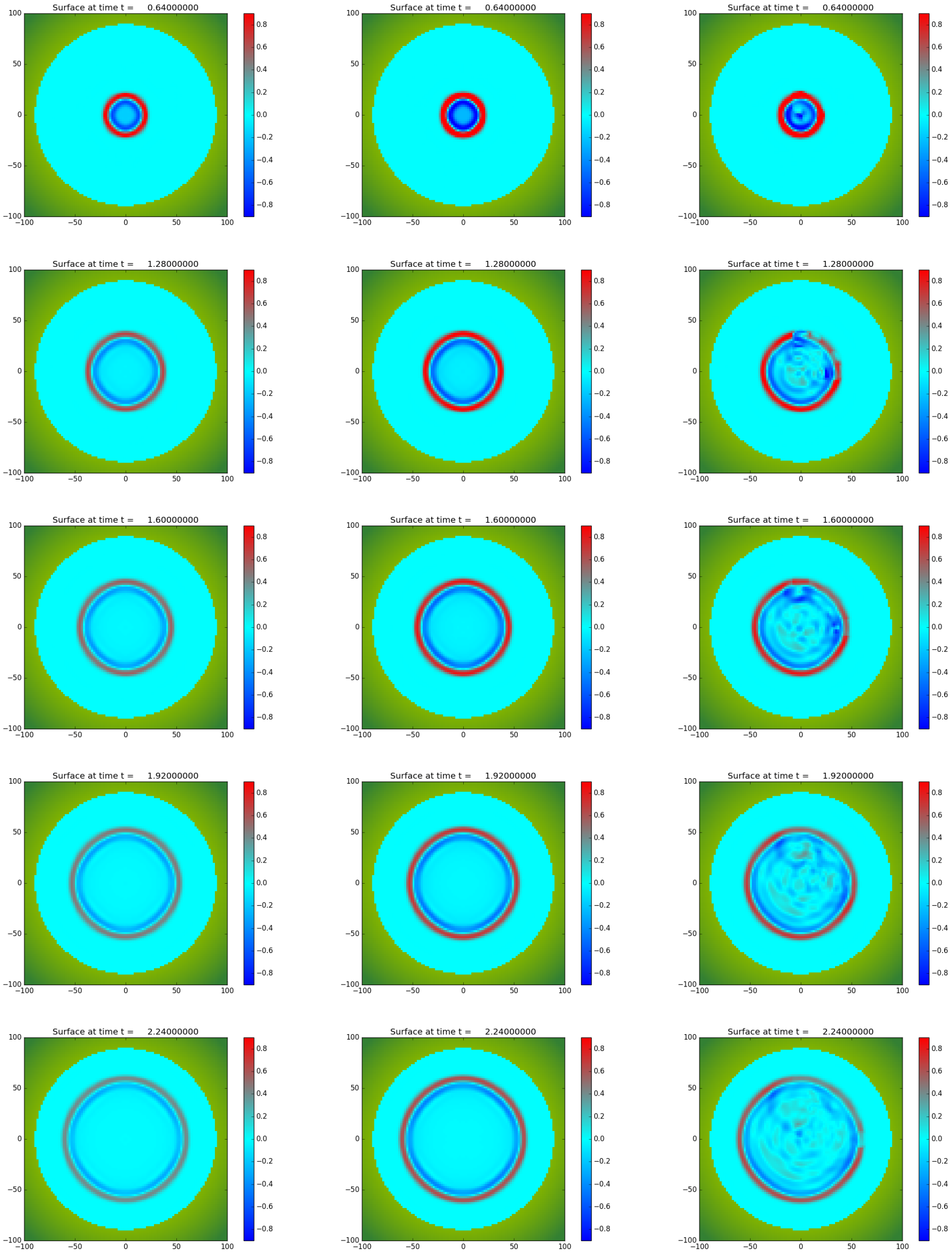


Figure 6.4: Error comparison between true solution, free run and the assimilated solution for testing effect of filter type on assimilation.

6.1.2 Assimilation and initial condition error

The objective of this test case is to assimilate the solution with a biased initial condition while using the observation values from the run that is considered as *truth*. The Gaussian hump in truth has an amplitude of 10 meters and the biased run has an amplitude of 15 meters. Figure 6.5 shows the *truth*, the *free run* and the assimilated solutions. As seen in Figure 6.5b, the traveling wave of the free run has higher amplitude as expected. Assimilation is performed using the observation configuration of Figure 6.1 with the values from the truth. The figure shows the result for a single level refinement of cells. The assimilation parameters are same as given in Table 6.2.

As seen from Figure 6.5c, the leading wave in the assimilated solution gets updated with the lower amplitude of observations values from the true solution. Further, Figure 6.6 demonstrates that the uncertainty in the water surface elevation of the leading wave reduces with time, as more number of assimilations occur.Note that, only the top right region gets updated which is consistent with the fact that local assimilation is being performed with observation values only in that region. This encourages to adopt assimilation for tsunami applications wherein the leading wave can be corrected in case of any errors discussed so far.



(a) Truth (b) Free run (c) Assimilated

Figure 6.5: Snapshots indicating the standard deviation of water surface elevation in the radial bowl test case to observe the reduction in error in the initial condition. Left - *Truth*; Center - Free run 2; Right - Assimilated

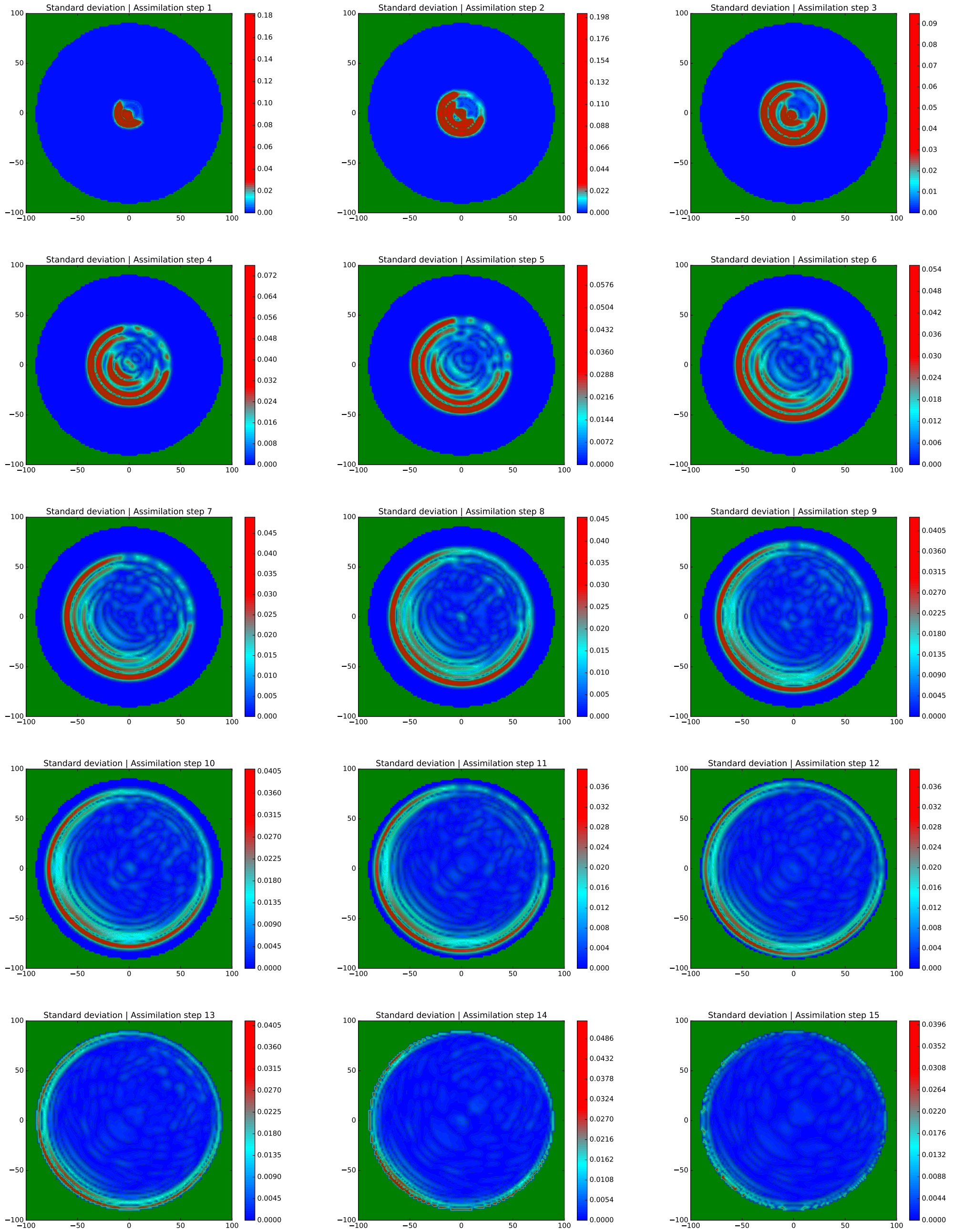


Figure 6.6: Snapshots indicating the standard deviation of water surface elevation in the radial bowl test case to observe the reduction in error in the initial condition. Left - *Truth*; Center - Free run 2; Right - Assimilated

6.1.3 Summary

Two test cases are setup to assess discretization errors and the initial condition errors.

The test setup for analyzing discretization error uses observation values from a simulation with highly refined mesh size and assimilated with the coarse mesh water surface elevation, while keeping the same initial condition. It is found that assimilation improves the solution on the coarse mesh and thus can decrease the discretization error. Further, it is found that localized SEIK version works the best in the current test case. However, it must be noted that other variables such as localization radius, observation rms etc. may affect the performance of a filter type. But overall, the result is that assimilation aids in error reduction.

The test setup for analyzing the errors in initial condition uses the same mesh size for truth and assimilation run. The true solution is generated using an initial condition and observation values are extracted as per the observation configuration. The assimilation run is performed with a bias with respect to this initial condition and the water surface elevation is assimilated. It is found that assimilation reduces the error between true solution and the free run. Further, it is found that the leading wave gets updated with respect to the truth. This is especially important for tsunami applications where there may be very high errors in the leading wave that is generated from the initial bathymetry changes that have high uncertainty.

6.2 Chile 2010 tsunami event

On February 27, 2010 at 03:34 local time (06:34 UTC), an 8.8 magnitude earthquake off the coast of Chile triggered a massive tsunami, causing catastrophic damage to life and property.

In this section, we will apply the discussed methodologies of initial ensemble generation and AMR coupled data assimilation framework on a test setup that simulates the event. We first describe the model parameters for simulating the actual tsunami event. To assess the assimilation system, we perform twin experiments, wherein we treat the above run as the "truth” and use the water surface elevation data as measurement values for the assimilation runs. We will demonstrate that assimilation is effective on model runs that have the initial conditions that are different from the true initial conditions. Note that the initial condition here refers to the fault parameters causing the associated earthquake. We will also perform experiments to assess the effect of localized assimilation and the placement of measurement sensors on the assimilated solution. This is followed by an additional twin experiment that analyzes the effect of initial ensemble generation and the ensemble size on assimilation runs. Further, we also validate our results using real gauge data of the actual event. In the process we will demonstrate the efficacy of assimilation with AMR to obtain improved state estimates and the corresponding uncertainty.

6.2.1 Model parameters

For the tsunami simulation, the computational domain is the South Pacific ocean off the Chilean coast shown in Figure 6.7. The domain contains islands/features such as Easter Island and the Galapagos Islands. In the coarsest mesh, these features are not visible. During the simulation, as the mesh refines, the features become more apparent. The model parameters, including the mesh parameters, time stepping parameters, and topography are given in Table 6.3. The marked points 15, 12 and 32412 represent locations where the quantities of interest, such as water surface elevation, are recorded. In Figure 6.7, the two black lines given by the horizontal line at $-30°$ and the vertical line at $-90°$, represent the partition of the domain at the coarsest grid level into 4 patches. This is related to the specifics of the GEOCLAW software.

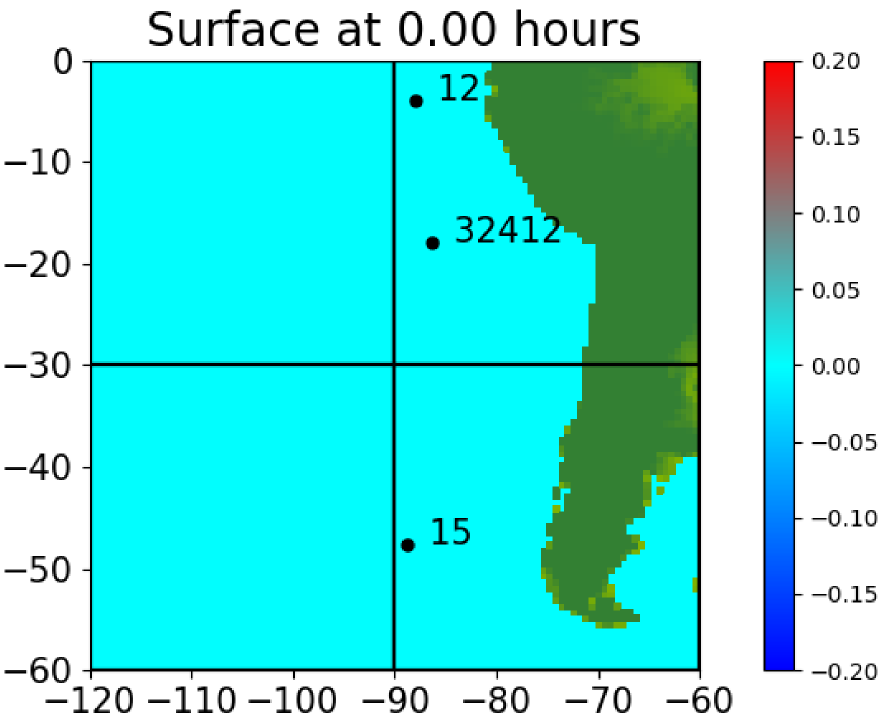


Figure 6.7: Computational domain for Chile February 27, 2010 tsunami simulation

Table 6.3: Model parameters for simulating Chile tsunami event of 2010

Computational domain	$x \in [-120, -60]$ and $y \in [-60, 0]$
Coarsest discretization	$100\ cells \times 100\ cells$
Simulation time	$13\ hours$
Time stepping	Initial time step $= 0.016\ seconds$ and desired $cfl = 0.75$
Topography	10 minute ETOPO2 from NGDC[1]

Table 6.4: Topography parameters for ensemble generation and twin experiment

Statistic	Strike	Length	Width	Depth	Slip	Rake	Dip	Longitude	Latitude
$Truth^*$	16.0	450.0	100.0	35.0	15.0	104.0	14.0	-72.668	-35.826
$Biased^*$	—”—	—”—	—”—	—”—	35.0	—”—	—”—	—”—	—”—

* Mean topography. Units $[meter, kilometer, kilometer, kilometer, meter, meter, meter,$

6.2.2 Reference solution

The reference solution is simulated using the “true" fault parameters of the Okada model, given in Table 6.4 that produces the associated slip in the fault region and the sea floor deformation shown in Figure 6.8a. The bathymetry change causes a tsunami with the water surface elevation snapshots given in Figure 6.9a. The figures show a primary wave traveling in the north-west direction. At approximately 4.0 hours, a section of the wave traveling along the north, reflects off the coast of Peru to form a secondary wave traveling in the south-west direction.

6.2.3 Twin experiment

In this experiment, we set up a model run with biased fault parameters in comparison to the truth. The twin experiment is aimed to demonstrate that the run

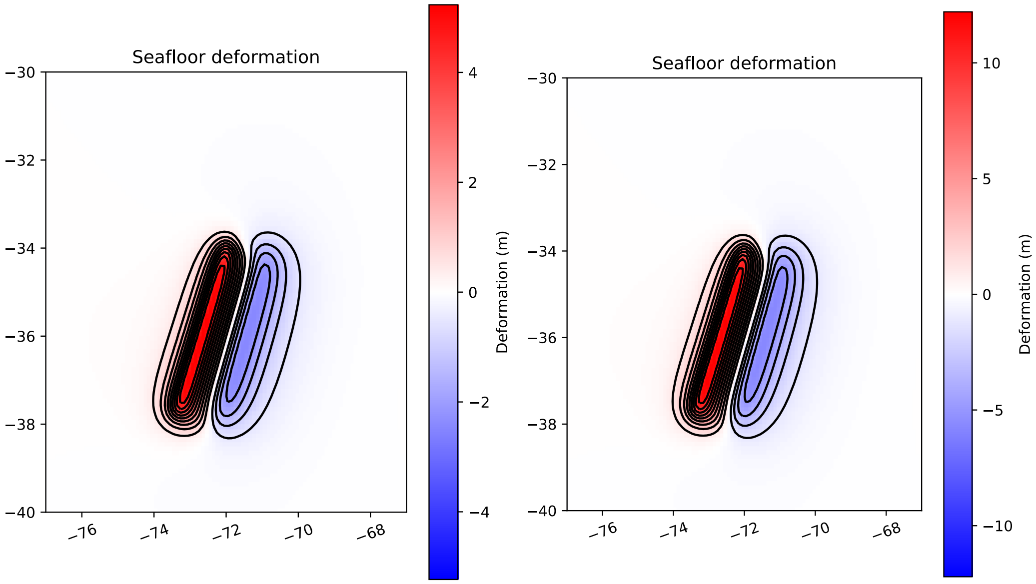
---

[1] National geophysical data center (NGDC) GEODAS grid translator 2010

with the biased fault parameters can be effectively assimilated using the measurements from the “true" solution.

6.2.3.1 Biased run

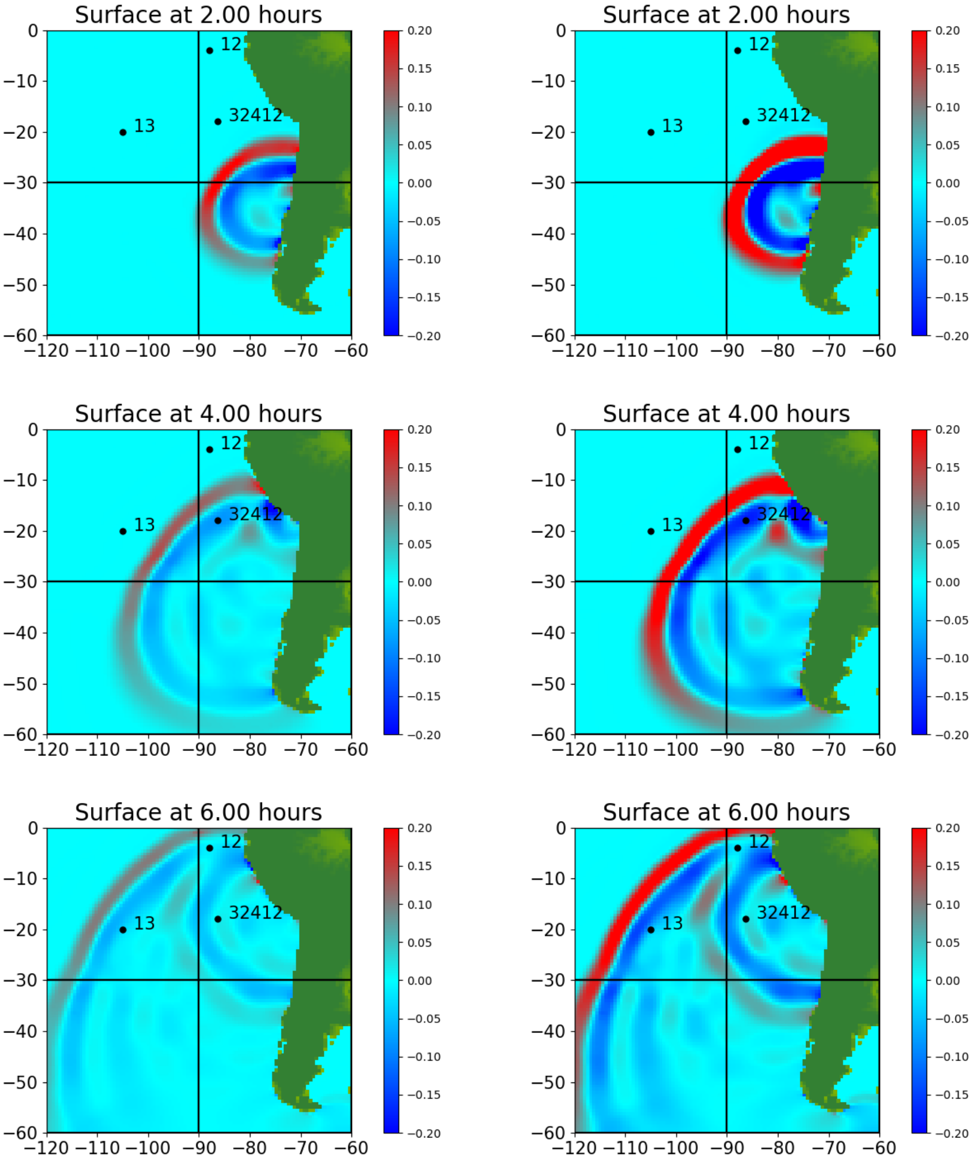
The bias in the fault parameters given in Table 6.4 is a slip of 35.0 meters compared to 15.0 meters for the truth. The additional slip in the fault regions relates to a larger bathymetry change shown in Figure 6.8b, leading to a stronger tsunami. This can be seen in the snapshots in Figure 6.9b for a free run under this biased initial condition. Note that free run means a model run that is not assimilated. Comparing the “true" state in Figure 6.9a with the free run state in Figure 6.9b, the tsunami generated from the biased fault parameters has a larger amplitude than the true state, till the waves exit the computational domain.



(a) “True" fault parameters

(b) Biased fault parameters

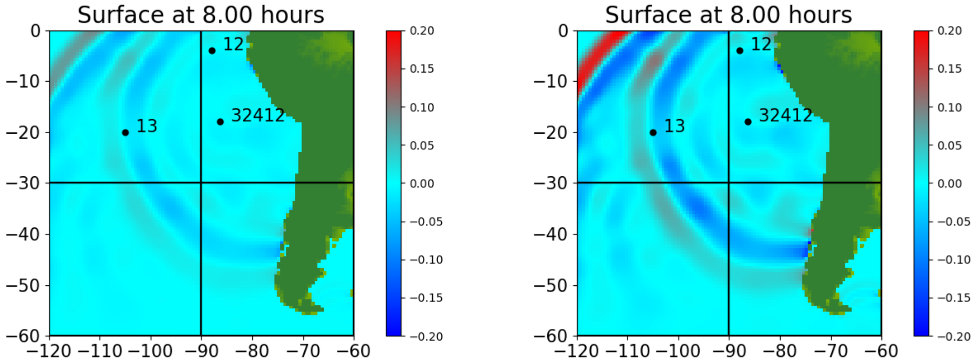
Figure 6.8: Bathymetry changes from the biased fault parameters in Table 6.4



(a) “True" fault parameters

(b) Biased fault parameters

Figure 6.9: Snapshots of water surface elevation at time 2.00 hours, 4.00 hours and 6.00 hours, caused by the tsunami produced from bathymetry changes due to the fault parameters in Table 6.4



(a) “True" fault parameters (b) Biased fault parameters

Figure 6.9, cont: Snapshots of water surface elevation at time 8.00 hours, caused by the tsunami produced from bathymetry changes due to the fault parameters in Table 6.4

6.2.3.2 Initial ensemble generation

To generate the initial ensemble of states, we implement the algorithm discussed in Section 5.4. We treat the biased slip fault parameter as the mean slip for the free run. We take a sample of 100 (N_s) slips assuming that the slip follows a normal distribution of mean 35.0 meters and a standard deviation of 4 meters. Figure 6.10 shows the sample distribution of the slips generated with a mean of 20.0 meters added to the true slip of 15.0 meters. A forward integration of the model until the checkpoint time (t_{cp}) of 2.0 hours, with each unique fault parameter as an initial condition, results in the subspace of tsunamis. Proceeding with a singular value decomposition of the covariance matrix, followed by the construction of the perturbation matrix, we generate the initial ensemble of desired size. For illustration, Figure 6.11 is the mean water surface elevation at the checkpoint time (t_{cp}) of 2.0 hours for the model run simulated using the biased fault parameter as the initial condition. Figure 6.12 shows the first 9 members of the 16 ensembles (n_s) that are generated after using the initial ensemble generation algorithm. At the checkpoint time (t_{cp}) of 2.0 hours, we resume the simulation, while assimilating the ensemble members with the measurement data from the *true* solution.

For all our experiments,we have applied ESTKF to assimilate the synthetic data every 15 minutes, setting the inflation factor to 1.1. We assess the assimilated water surface elevation, in comparison to the model free run (non-assimilative run). The error norm used is,

$$\|e(t)\|_2 = \|\psi_{true}(t) - \hat{\psi}(t)\|_2,$$

where $\hat{\psi}$ can be the assimilated state or the free run state.

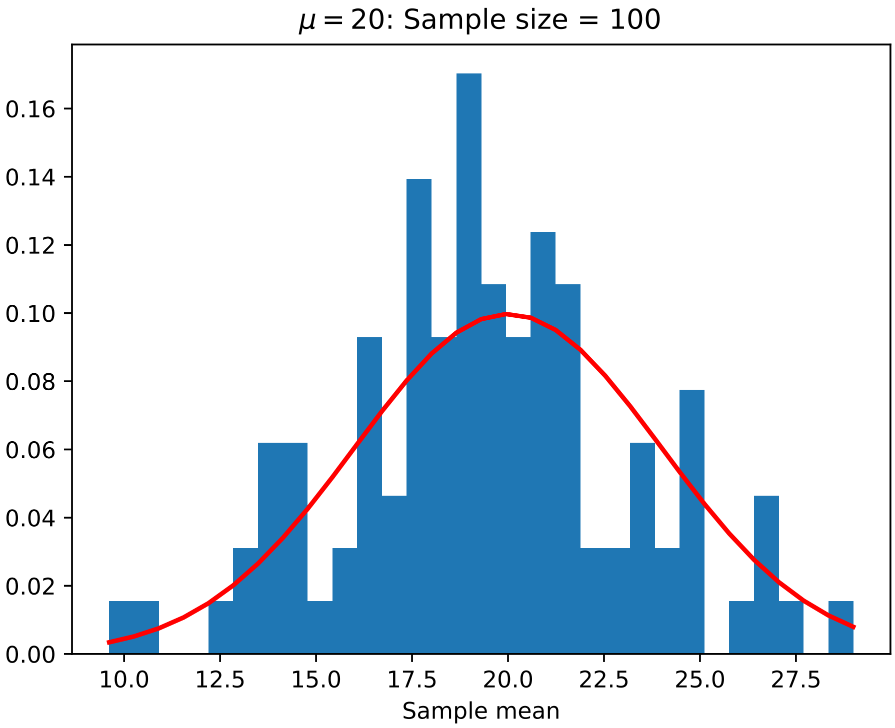


Figure 6.10: Sample distribution of slips for the initial ensemble generation algorithm

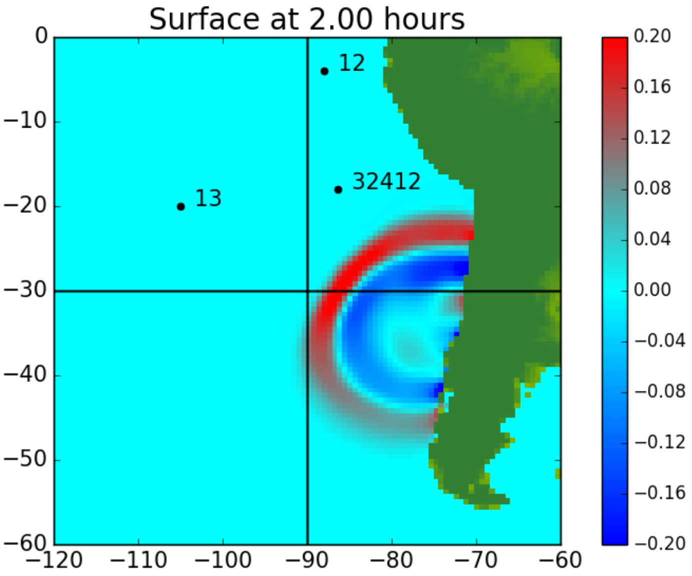


Figure 6.11: Water surface elevation at the checkpoint time of 2 hours.

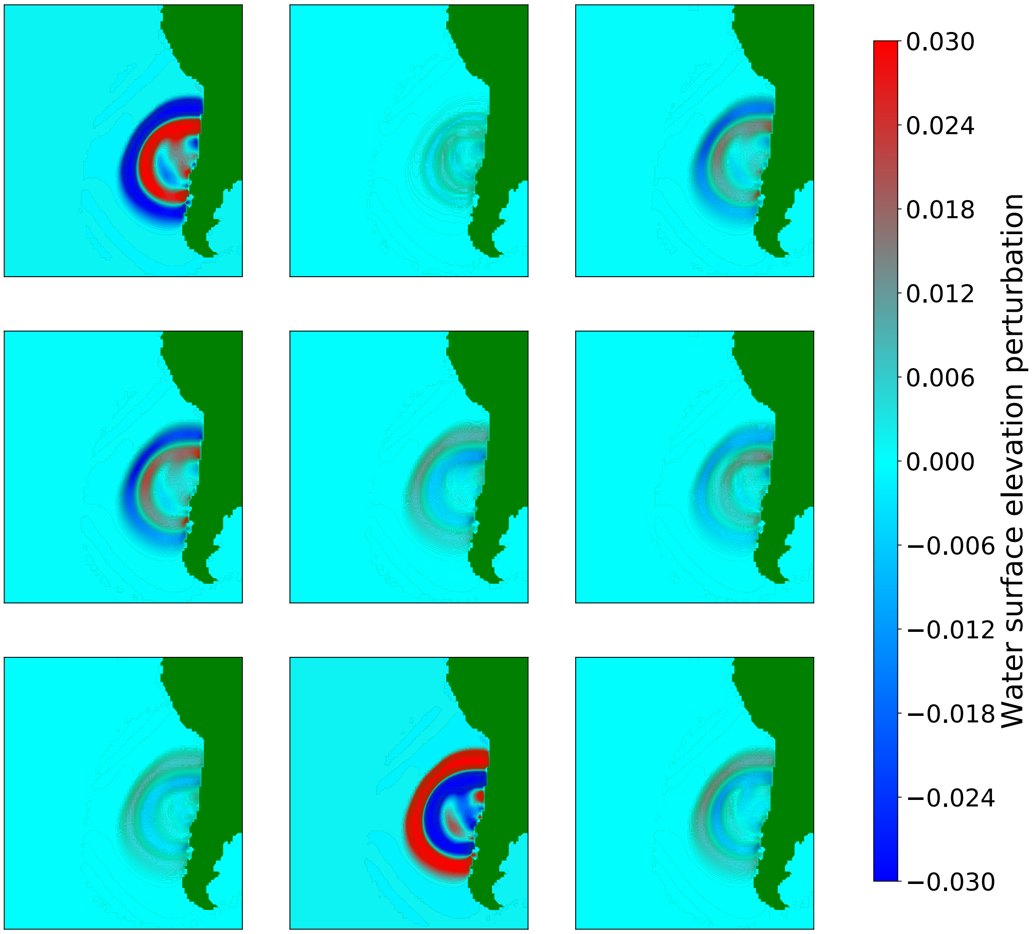


Figure 6.12: Perturbations in water surface elevation that will be added to the mean state at the checkpoint time $t_{cp} = 2.0$ hours

6.2.3.3 Experiment 1: Effect of observation configuration and localization radius

In this experiment, we analyze the effect of the number of measurements and localization radius on the assimilated state estimates. We have chosen various observation grids ranging from the region containing 4 (a sparse network), 12°, 35°

and 42° (a dense network) sensors, distributed in the Pacific ocean as shown in Figure 6.13. For each of the observation grids, we use 16 (n_s) ensemble members with localization radius of 3°, 5°, 10°, 15°, 20°, 25° and 35°, along with a global assimilation. The observation error is assumed to be normally distributed with the standard deviation of observation error set to 0.005 meters. At every assimilation step, the state estimate after the assimilation is compared to the truth and the errors are calculated. Note that the *true* state is given in Figure 6.9a and the free run is given in Figure 6.9b.

At the fault region, varying slip parameters for the truth and the biased model runs leads to a differing bathymetry profile. For error calculation for this experiment, we omit the region around the fault, $(-76.0^\circ, -68.0^\circ) \times (-40.0^\circ, -32.0^\circ)$. The error plots for all the mentioned observation configurations are given in Figure 6.14. It was found that the error plots produced without ignoring the fault region give a similar trend as the plots in Figure 6.14. However, ignoring the fault region makes the trend more pronounced. The figures plot the difference between the free run and the truth, the difference between the globally assimilated state estimate and the truth, and the difference between the locally assimilated state estimate and the truth for various localization radii.

Figure 6.14 indicates that irrespective of the observation configuration, as the localization radius increases, the estimated error between the “true" solution and the assimilated state decreases till a *critical localization radius*. Assimilation with further increase in the localization radius (continuing towards a global assimilation) results in errors between assimilation run and the truth being greater than the error between free run and the truth, implying that assimilation starts to degrade the solution and performs worse than the case when the state was not being assimilated. Also, even with a small ensemble of 16 members (n_s), for the assimilation runs with localization radius less than the critical radius, the assimilation run estimates have smaller error compared to the error between non-assimilative free run and the truth, until at least approximately 30 assimilation steps, equal to 450 minutes (or 7.5 hours) of simulation. This corresponds to the time period of 2 hours to 9.5 hours, which may be critical for evacuation measures. For instance, the preparation time required for incoming tsunami at Hawaii is around 10 hours when the tsunami is generated from Chile and 4 hours when the tsunamis is generated from Alaska[2]. Hence, assimilation proves to be effective during this critical time by providing improved estimate of the state.

It is found that after 10 hours of simulation, the error norms are practically the same, whether data are assimilated or not. That is the assimilated run converges to the free run. This can be clearly observed in the assimilation of 42 observations in Figure 6.14d. This is because, after around 10 hours, the tsunami wave front has traveled across the computational domain, leaving behind no significant wave fronts in the region. However, in assimilation of other observation grids, the errors seem to increase after nearly 30 assimilation steps. This is attributed to multiple factors. One of the reasons is that a smaller localization radius does not cover the regions near the coast and some small errors persist. The assimilation with 42 observation grids witnessed local assimilation of the coastal regions as well and this reduced the

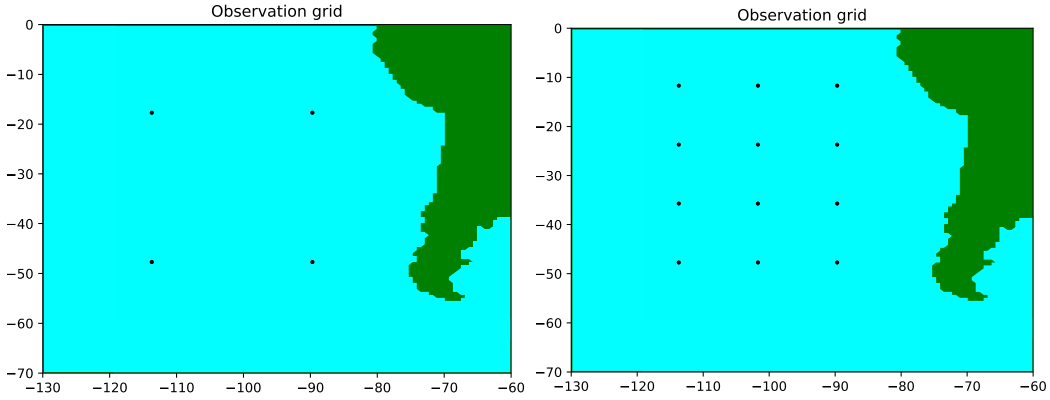
---

[2] http://ptwc.weather.gov/hawaii.php

error further. Thus, for this setup, it is found that “small" localization radius with measurements closer to the coastal areas can prove to be effective for assimilation. Increased errors near the end of the simulation can also be attributed to errors due to small ensemble size and ensemble variance inflation at every assimilation step. In Section 6.2.3.4, we will demonstrate that by taking larger ensemble size, this error can further be reduced.

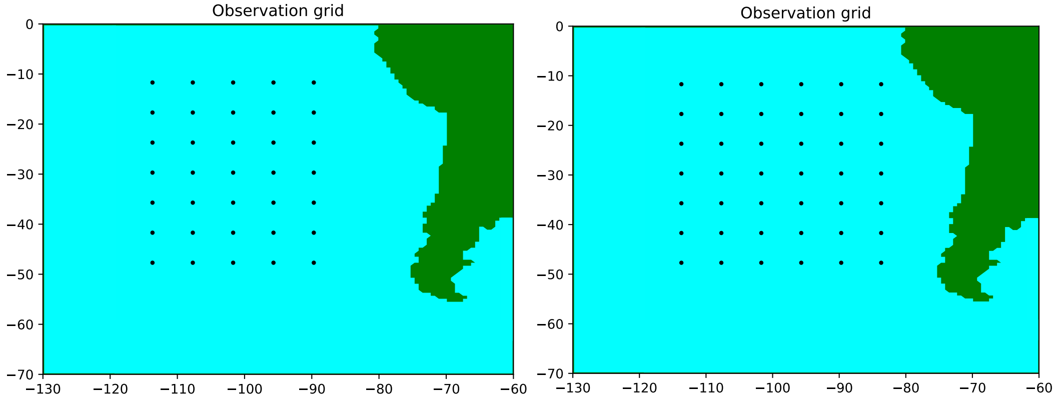
For the observation configuration with 4, 12, 35 and 42 observations, the critical localization radius is approximately $25°$ (approximately 2500 kilometers of land), $15°$ (approximately 1500 kilometers of land), $15°$ and $15°$, respectively. The state estimate for each of the assimilation runs with the critical localization radius is given in Figure 6.15a, Figure 6.15b, Figure 6.15c and Figure 6.15d respectively. From the state estimates with assimilation using observation configuration of 4 sensors and 12 sensors (Figure 6.15a and Figure 6.15b), at 4.0 and 5.0 hours, the tsunami wavefront has smaller estimated water surface elevation (amplitude) when assimilation is performed with 12 observations, when compared to assimilation with 4 observations. The smaller amplitude conforms to the truth at the time of 4.0 and 5.0 hours in Figure 6.9a. This suggests that in this particular experiment, assimilating with 12 observations is more effective than using 4 observations. We further compare the filter state estimate (Figure 6.15b and Figure 6.15c) obtained by assimilating observations with configuration of 12 sensors to the filter state estimate obtained by assimilating observations with the configuration of 35 sensors. The observation configuration with 35 sensors has more sensors within the same area compared to the observation configuration of 12 sensors. Further, the edge that connects the rightmost sensors is equally away from the coast for both the configurations. It is observed that there is not much improvement in the estimated state, even though more observations are assimilated. The observation configuration with 42 sensors has 7 additional sensors to the right, compared to the 35 observation configuration. That is, the edge joining the rightmost sensors for observation configuration with 42 sensors are closer to the coast compared to the observation configuration of 35 sensors. Comparing water surface elevation snapshots at around time of $4.0, 5.0$ hours in Figure 6.15b and Figure 6.15c, assimilation using 42 sensors results in waves with higher estimated water surface elevation traveling along the coast when compared to assimilation with 35 observations. This reaffirms the fact that assimilation of sensor data closer to the coast may be helpful.

In summary, this experiment illustrates that the assimilation runs do not always guarantee that the analysis state will have lesser errors than the free run solution. In fact, in this experiment, the assimilation runs with high localization radius perform worse than the free run. One of the possible reasons might be that with high localization radius, the analysis step updates the state vector unreasonably in the regions that are physically far from the observation coordinates. Other reasons include existence of various confounding variables such as ensemble size, observation configuration, standard deviation of observation standard error and most importantly the difference of the initial condition and the truth. To get insights into this phenomenon, we have analyzed assimilation run with various localization radii, while keeping the above-mentioned variables fixed. It is found there exists an optimum localization radius that guarantees that the assimilation is effective. Thus, this experiment of assessing the optimum localization radius serves as a method for model calibration.



(a) 4 measurement sensors

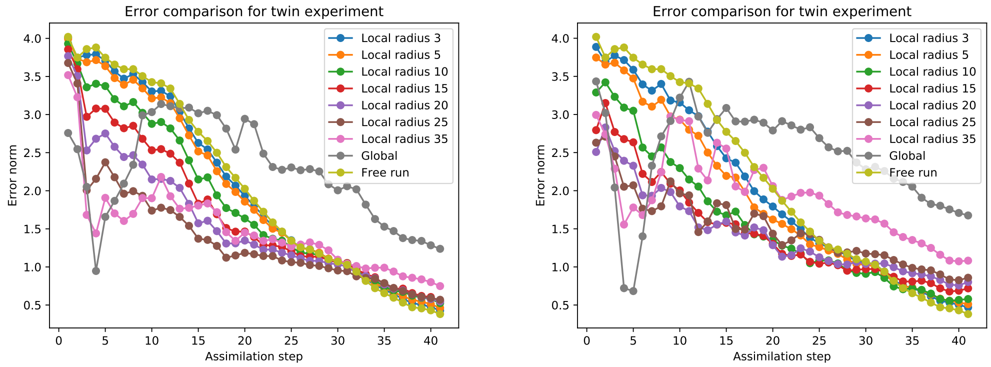
(b) 12 measurement sensors



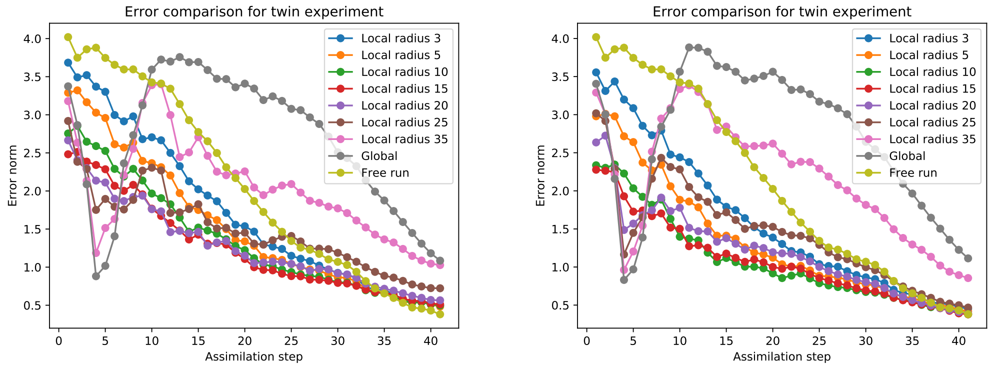
(c) 35 measurement sensors

(d) 42 measurement sensors

Figure 6.13: Observation grid configurations for twin experiment 1

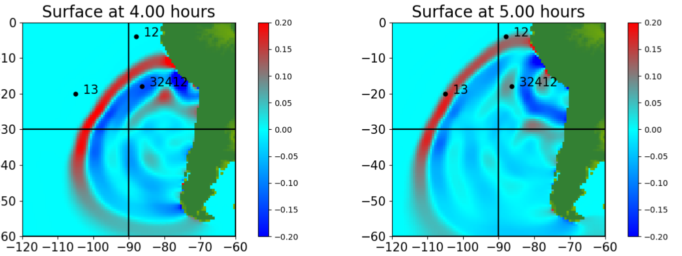


(a) Observation configuration of 4 sensors (b) Observation configuration of 12 sensors

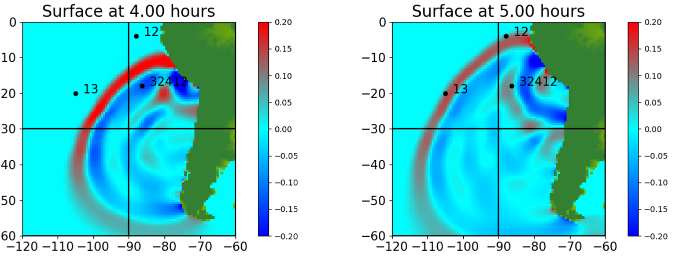


(c) Observation configuration of 35 sensors (d) Observation configuration of 42 sensors

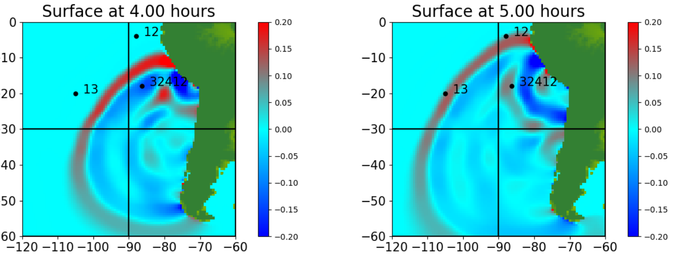
Figure 6.14: Error comparison for twin experiment 1 with various observation configurations given in Figure 6.13



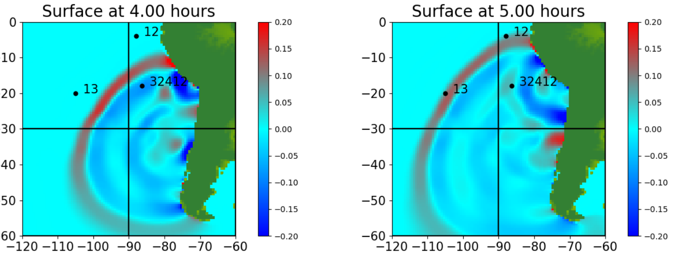
(a) 4 synthetic observations with the localization radius of 25°



(b) 12 synthetic observations with the localization radius of 15°



(c) 35 synthetic observationswith the localization radius of 15°



(d) 42 synthetic observations with the localization radius of 15°

Figure 6.15: Snapshots of mean water surface elevation in twin experiment 1 of Chile 2010 tsunami after synthetic observations are assimilated every 15 minutes.

6.2.3.4 Experiment 2: Effect of initial ensemble generation

In the previous experiment, the initial ensemble was generated via perturbations in the slip parameter only. Further, though the prior slip information of 35.0 meters was significantly away from the true slip of 15.0 meters, the above experiment demonstrated the effectiveness of assimilation. This experiment differs in a few aspects - 1) the initial ensemble is generated with perturbations in all the fault parameters, 2) the observation grid chosen is 142 sensors in Figure 6.16, covering regions that are closer to the shore, 3) the synthetic observations are generated by a model run that takes a random sample of fault parameters from the multivariate Gaussian distribution assuming the mean of the fault parameters is given by μ_{topo} in Table 6.5 and covariance matrix is diagonal with the diagonal entries given by σ_1^2, 4) the sample of 100 (N_s) fault configurations for the initial ensemble generation algorithm assumes a distribution of fault parameters with mean parameters given by μ_{topo} in Table 6.5 and variance terms of the parameters given by σ_2^2, 5) the observation error is assumed to be normally distributed with the standard deviation of observation error set to 0.001 m.

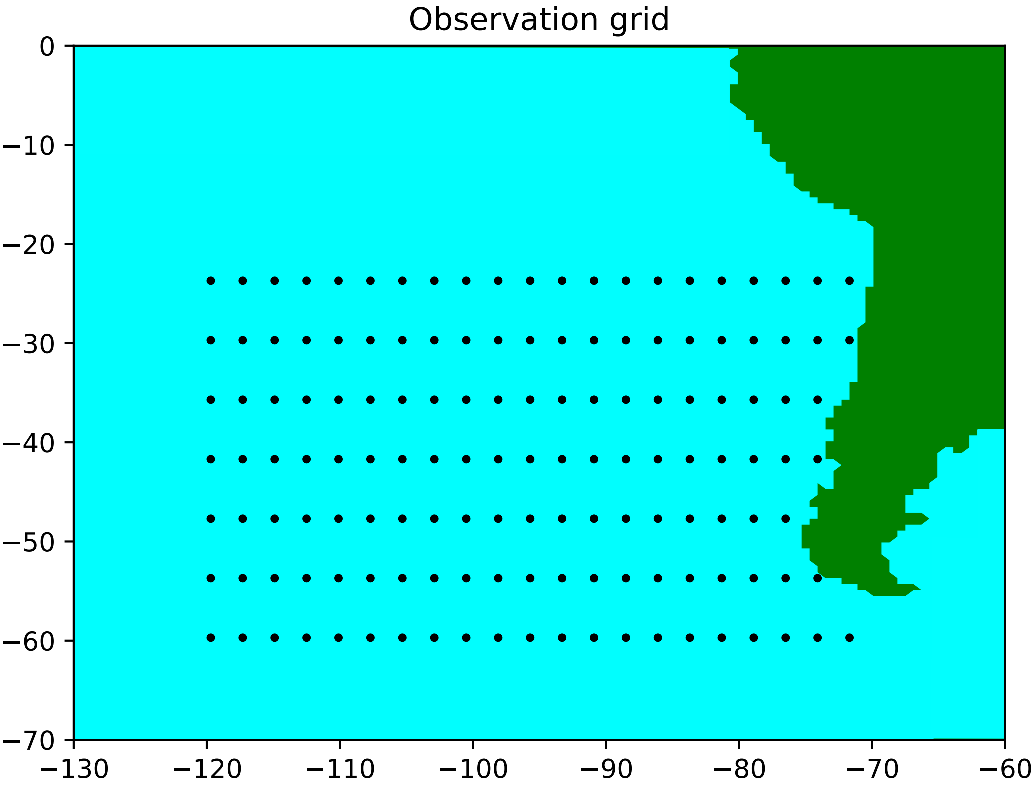


Figure 6.16: 142 observation points where synthetic measurements are generated for the twin experiment 2 of Chile tsunami simulation

Table 6.5: Topography parameters for ensemble generation for twin experiment 2

Statistic	Strike	Length	Width	Depth	Slip	Rake	Dip	Longitude	Latitude
μ_{topo} [a]	16.0	450.0	100.0	35.0	15.0	104.0	14.0	-72.668	-35.826
σ_1^2 [b]	4.0	150.0	50.0	10.0	5.0	20.0	4.0	3.0	3.0
σ_2^2 [c]	1.6	45.0	10.0	3.5	1.5	10.4	1.4	2.0	2.0

[a] Mean topography. Units $[meter, kilometer, kilometer, kilometer, meter, meter, meter, ^\circ, ^\circ]$

[b] Variance for generating initial ensemble for twin experiment ($Units^2$)

[c] Variance for generating synthetic observations ($Units^2$)

We can analyze the effect of perturbing all the fault parameters in the process of generating the initial ensemble mentioned in Section 5.4. Figure 6.17 plots the eigen decomposition of the covariance matrix of 100 (N_s) model states, with the red line indicating the cumulative percentage variance explained by the eigen modes. From the figure, it can be seen that the first 10 eigenvalues explain more than 85% of the total variance. We can also analyze the effect of the sample size on the percentage variance explained by the eigen modes in Table 6.6. The table suggests that as the sample size increases, a higher number of eigen modes account for the unique variance among the states. It is found that the subspace of the initial ensemble that can be generated is richer when compared to just perturbing the slip as in twin experiment 1.

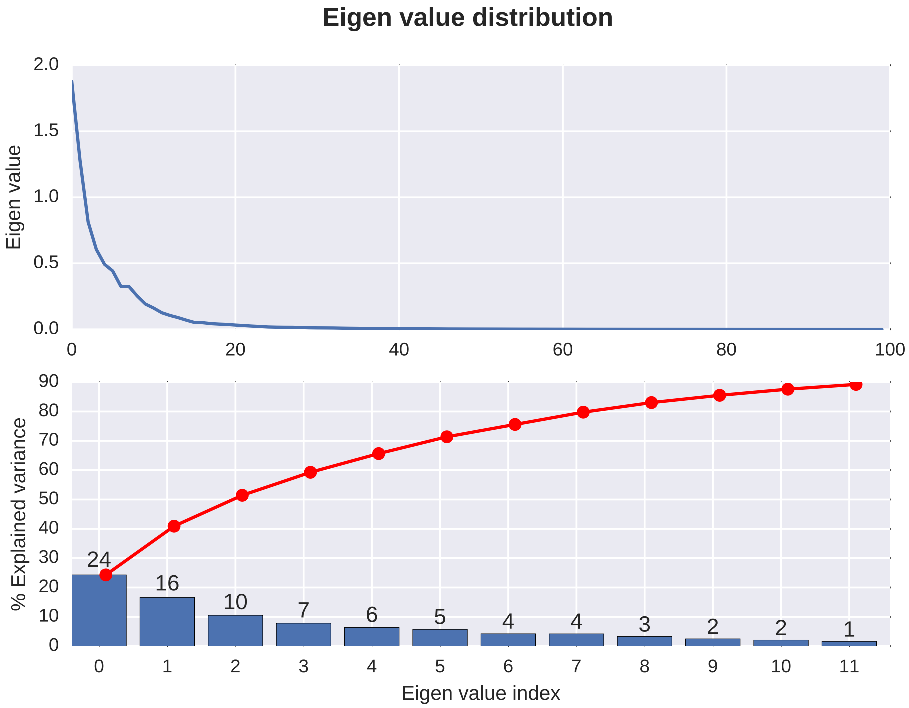
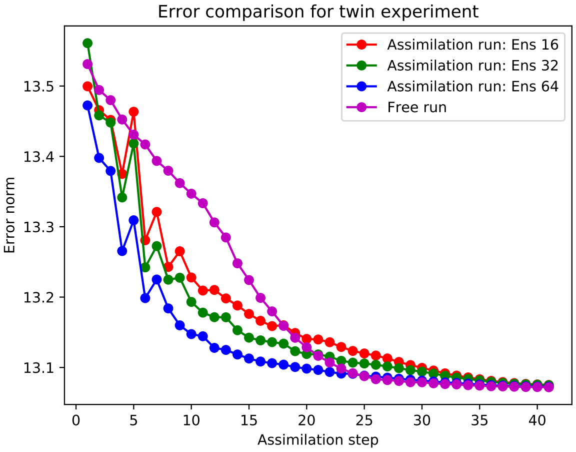


Figure 6.17: Eigen decomposition of 100 states of the possible tsunami outcomes at the checkpoint time of 2.0 hours for twin experiment 2 of February 27 2010 Chile tsunami data assimilation.

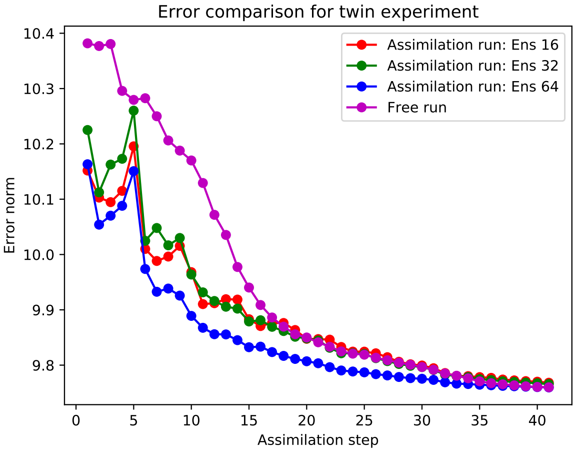
Table 6.6: Percentage captured variance with varying ensemble size in initial ensemble generation in twin experiment 2 of the Chile tsunami simulation

Ensemble	% explained variance by eigen modes												
	1	2	3	4	5	6	7	8	9	10	11	12	Total
16	32	21	12	10	7	5	3	3	1	0	0	0	100
32	27	18	11	9	7	6	4	3	2	2	1	1	91
40	27	19	10	8	7	6	4	3	2	2	1	1	90
64	24	18	10	8	6	6	4	3	2	2	1	1	85
80	24	17	10	7	6	5	4	4	3	2	2	1	85
100	24	16	10	7	6	5	4	4	3	2	2	1	84

Since we have taken a reasonably dense observation network, with observations covering near the coastal regions, the error calculations need not ignore the fault regions. Moreover, it is found that lower observation standard deviation of 0.001 meters gives reasonably good performance for the global assimilation as well. However, this is subjective to various factors defining the test setup such as the initial conditions, observation network etc. Figure 6.18a plots the difference between the “true" state and free run (non-assimilative run) state, and the difference between the “true" state and the globally assimilated state estimate for various ensemble sizes at the assimilation times. By analyzing the effect of ensemble size, it is found that an assimilation run with 64 members consistently produces more accurate state estimates than the non-assimilative free run. Further, as the ensemble size increases, the error decreases. It is found that the number of members required depends on how far the initial ensemble is from the truth. That is, by repeating the experiment with different sets of synthetic observations, it is found that in some cases assimilation run with just 32 ensemble members are sufficient to produce consistently less error than the free run. This is tested by another set of synthetic measurements obtained from a model run that uses the topography parameters sampled from a multivariate normal distribution with mean and variance as μ_{topo} and σ_2^2 respectively in Table 6.5. Assimilation is performed using the same initial ensemble that was generated earlier. From Figure 6.18b, assimilating the new measurement values results with 32 ensemble members providing consistently smaller errors than the free run. This suggests that the minimum ensemble size required for effective assimilation depends upon how far the first guess of the initial condition is from the truth.



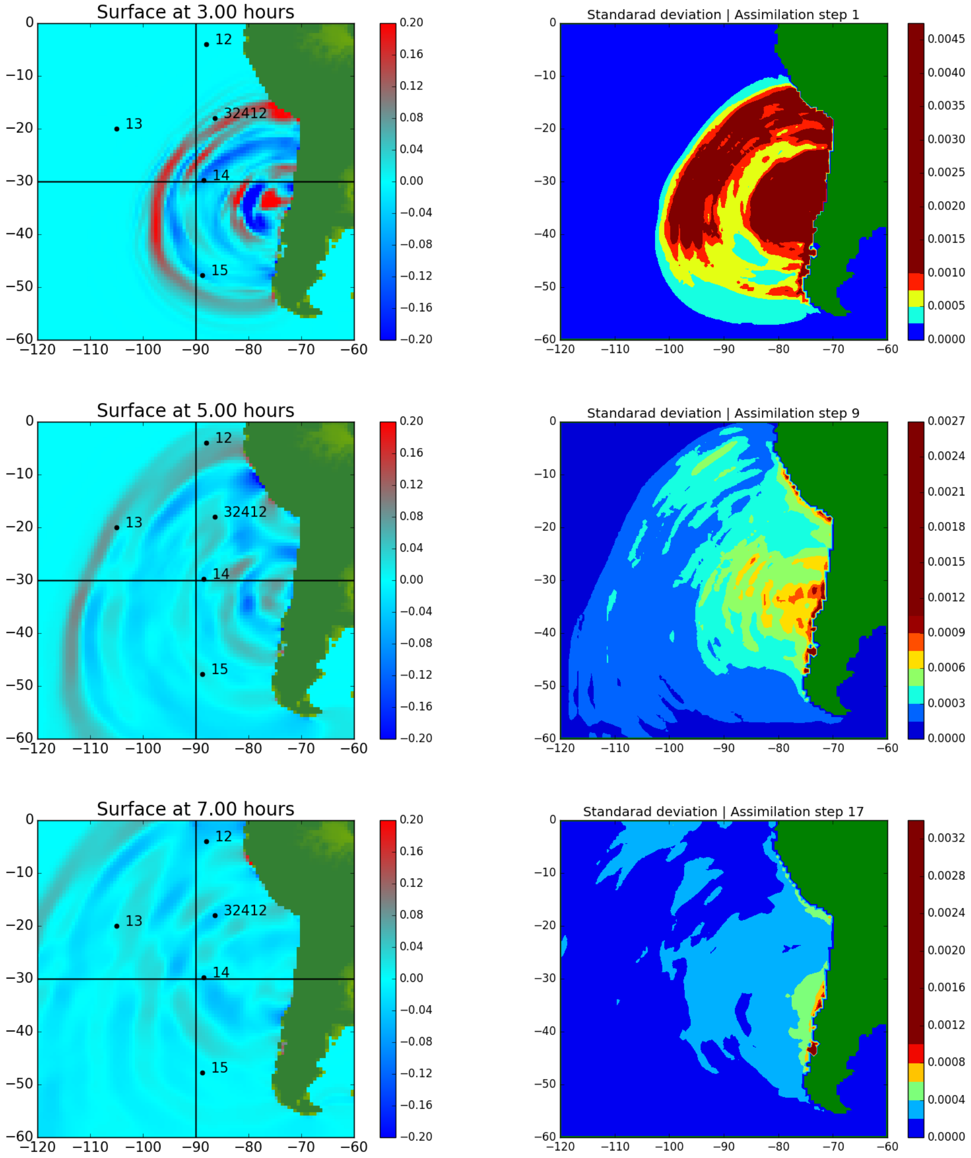
(a) Assimilation demonstrating use of 64 ensemble members to obtain errors between assimilated state and true state consistently less than errors between free run state and true state for every assimilation step



(b) Assimilation demonstrating use of 32 ensemble members to obtain errors between assimilated state and true state consistently less than errors between free run state and true state for every assimilation step

Figure 6.18: Effect of ensemble size on errors between true state and the global assimilated state with different sets of measurement values obtained from ensemble generation algorithm involving sampling of fault parameters as per Table 6.5.

For the case just described (Figure 6.18b), Figure 6.19a shows the estimated state at various assimilation steps and Figure 6.19b shows that over the simulation time, successive assimilations reduce the root mean square error in the water surface elevation. This translates to the fact that with every assimilation step, we obtain state estimates with better certainty. Also, some uncertainty persists near the fault regions that decreases with further assimilation steps.



(a) Mean state (b) RMSE

Figure 6.19: Snapshots of mean and RMSE of water surface elevation in twin experiment of Chile 2010 tsunami after 142 synthetic observations are assimilated every 15 minutes for twin experiment 2

6.2.4 Real gauge experiment

The DART buoys measure the pressure at the sea floor in deep water to provide an early indication of the magnitude of a tsunami. Gauge data of the real Chile 2010 event is available via DART (Deep-ocean Assessment and Reporting of Tsunamis) buoy 32412[3]. For this experiment, we will illustrate data assimilation using 2 levels. Also, we will use observation localization with support radius $20°(\approx 2000km)$.

The standard deviation of observation error is set to 0.005 meters. Assimilation is done every 15 minutes with 16 (n_s) ensemble members and the inflation factor is set to 0.9. Figure 6.20 compares the assimilation mean state (right side) to the free run state (left side). The rectangular boxes in the snapshot represent refined regions by a factor of 2 in both directions. The figures show that the amplitude of the tsunami wavefront propagating in the west direction is reduced when the gauge data is taken into account. Also, since the assimilation is done on a supermesh, there are more level 2 regions in the assimilated state than the corresponding free run state. This is expected because at the assimilation step, refinement is performed in regions where at least one of the ensemble members has a level 2 region at the end of its forecast step. At time of 5.0 hours, unlike the free run, the best estimate of the water surface elevation snapshot for assimilation run reveals a series of waves traveling in south-east direction towards the coast. This may have implications that data

---

[3] `http://www.ndbc.noaa.gov/dart.shtml`

assimilation is effective at capturing waves that might otherwise have not been resolved in non-assimilative simulations. Further, the AMR capability captures these waves and refines along the regions until it reaches the coast. We observe that during the entire simulation, more refined regions are automatically obtained along the coast when assimilation is performed, in comparison to the non-assimilative free run. Thus, the AMR is effective in capturing the oscillations created by the state update. Most importantly, Figure 6.21 outlines the effect of assimilation through the gauge output for the free run and simulation with assimilation. With assimilation, the gauge output (bottom) closely follows the real observation data. This was not possible otherwise.

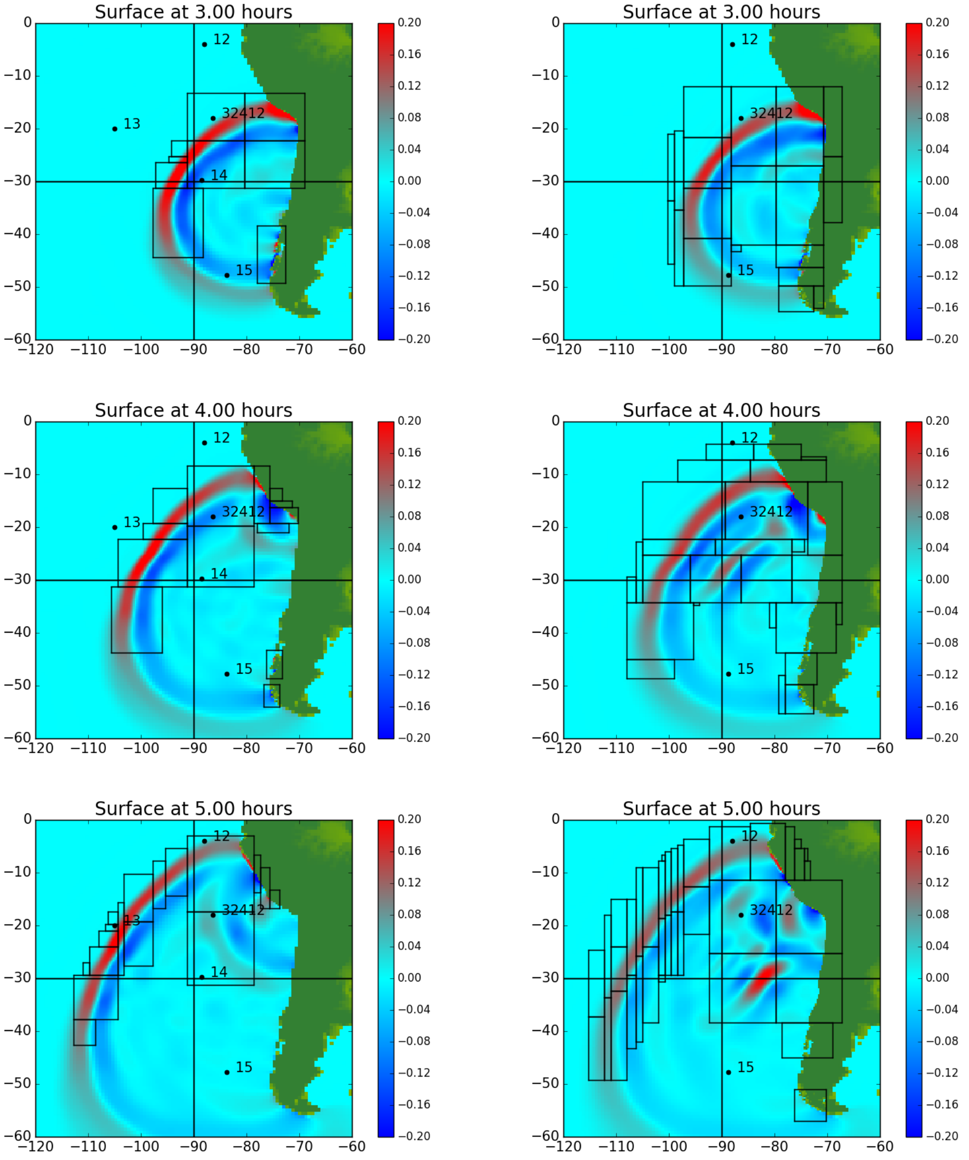


Figure 6.20: Water surface elevation comparison of (Right) mean assimilated state and (Left) state without assimilation after data from Gauge 32412 is assimilated every 15 minutes, with a maximum 2 level AMR at time 3.00 hours, 4.00 hours and 5.00 hours

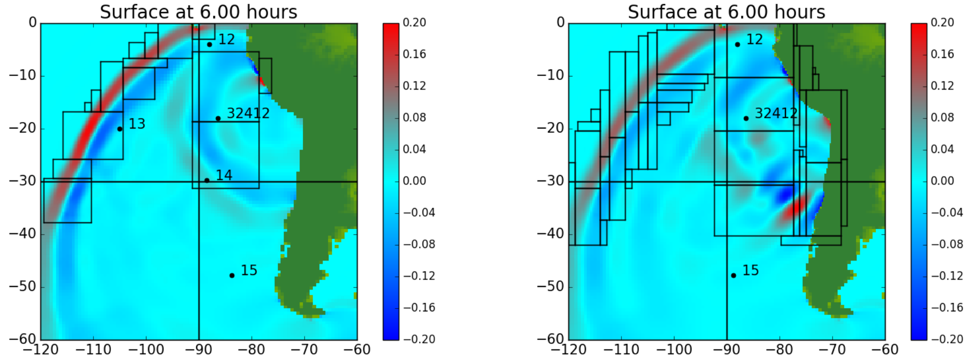


Figure 6.20, cont: Water surface elevation comparison of (Right) mean assimilated state and (Left) state without assimilation after data from Gauge 32412 is assimilated every 15 minutes, with a maximum 2 level AMR at time 6.00 hours

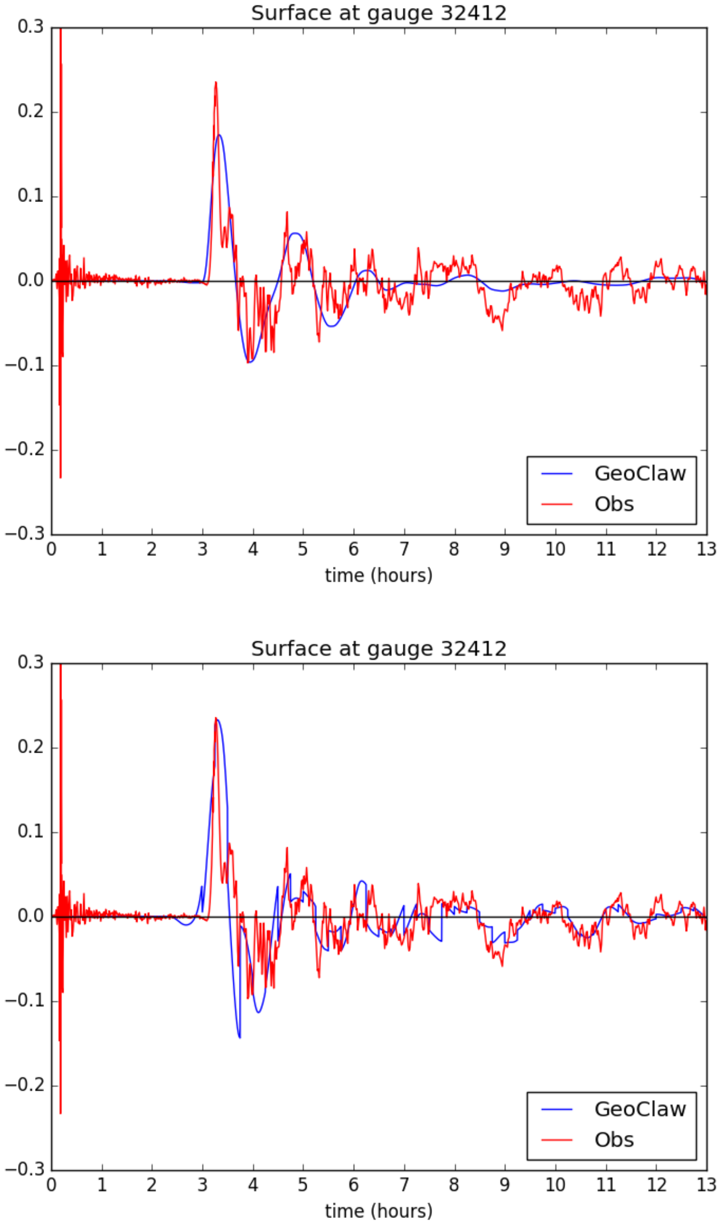


Figure 6.21: (Top) Gauge output for run without assimilation (Bottom) Gauge output for assimilation run of 16 ensemble members with localization radius of 20°

6.3 Tohoku tsunami event

The Tohoku tsunami event test case is the final test case where the lessons from the Chile test case on ensemble generation, localization etc are incorporated. The Tohoku tsunami test case simulates the tsunami off the coast of Japan.

The objective is to validate that data assimilation can improve forecasting, while taking into account the underlying uncertainties. Through the bathymetry parameters that are available during the event, it is possible to generate the seafloor deformation and hence a reference solution. Additionally, we have access to the actual water surface elevation measurements at multiple gauge locations. First the reference solution is compared to the gauge solution and the deficiencies are pointed out as given in Section 6.3.2. Next, similar to the Chile test case, Section 6.3.3 demonstrates the implementation of data assimilation by assimilating the gauge measurements. Additionally, the section illustrates the findings and draws conclusions and remarks.

6.3.1 Model parameters

Figure 6.22 shows the computational domain of Tohoku test case. The numbers marked in the figure indicate the locations where gauge data is available. However it should be noted that during the tsunami event, not all gauges are active for the entire period of the event. Table 6.7 defines the simulation parameters. The parameters defining the topography changes are given in Table 6.8.

Table 6.7: Model parameters for simulating Tohoku tsunami event

Computational domain	$x \in [130^\circ, 240^\circ]$ and $y \in [10^\circ, 60^\circ]$
Coarsest discretization	$55\ cells \times 25\ cells$
Simulation time	$11.66\ hours$
Time stepping	Initial time step $= 0.016\ seconds$ and desired $cfl = 0.75$

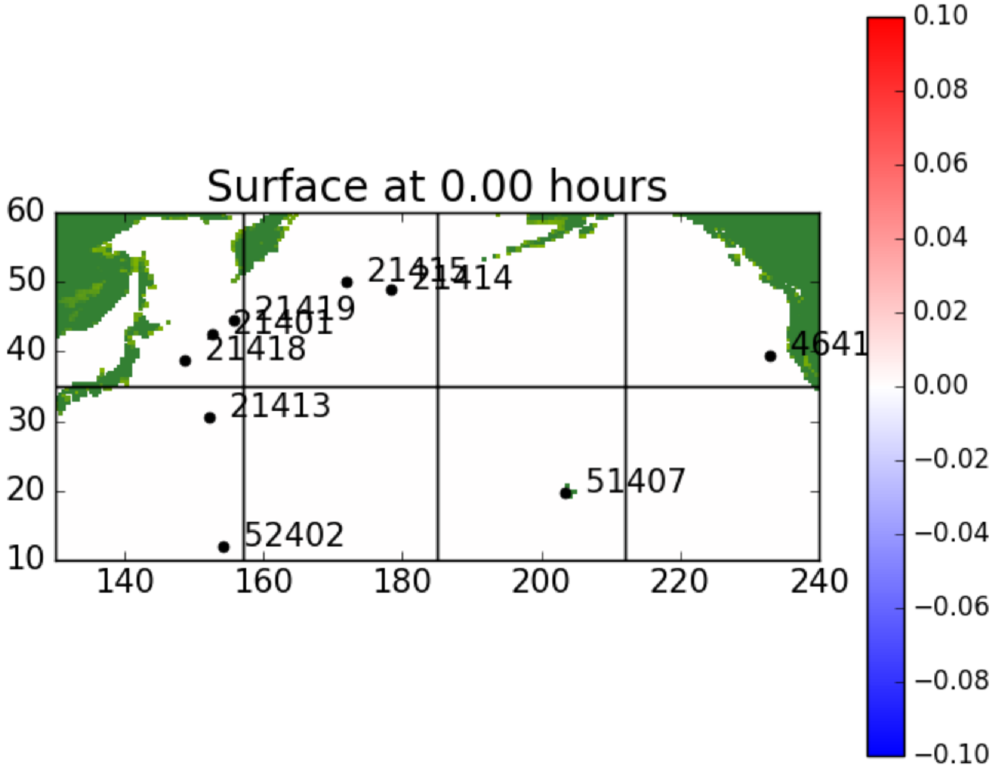


Figure 6.22: Computational domain for Tohoku tsunami test case

Table 6.8: Topography parameters for the base subfault during Tohoku tsunami event

Strike	Length	Width	Depth	Slip	Rake	Dip	Longitude	Latitude
198.0	475.0	200.0	7.5052	Figure 6.23	90.0	10.0	37.64165	143.72745

* Units $[meter, kilometer, kilometer, kilometer, meter, meter, meter, ^\circ, ^\circ]$

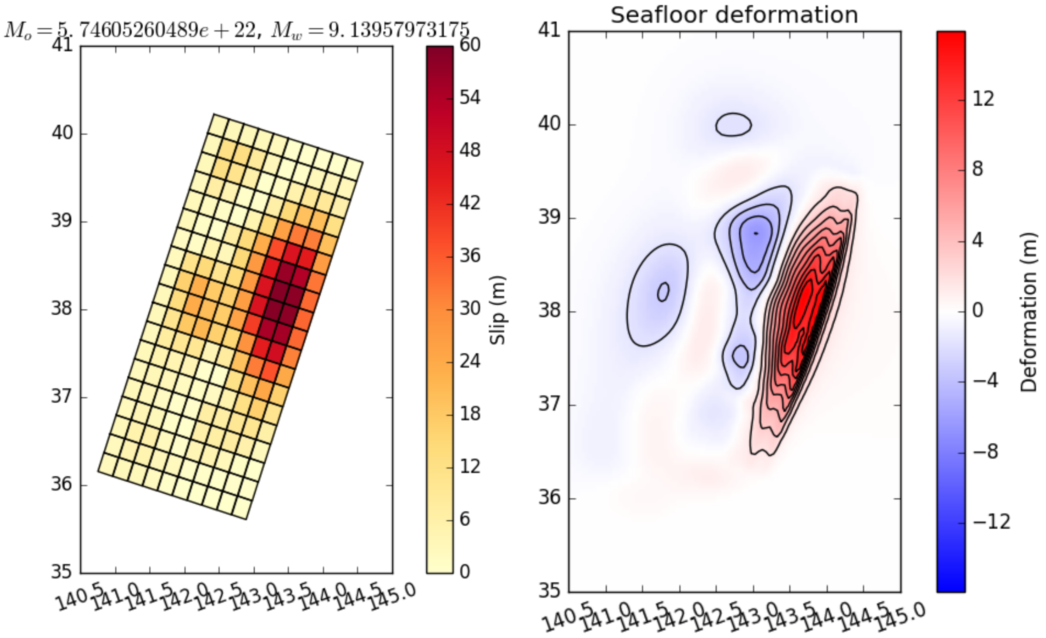
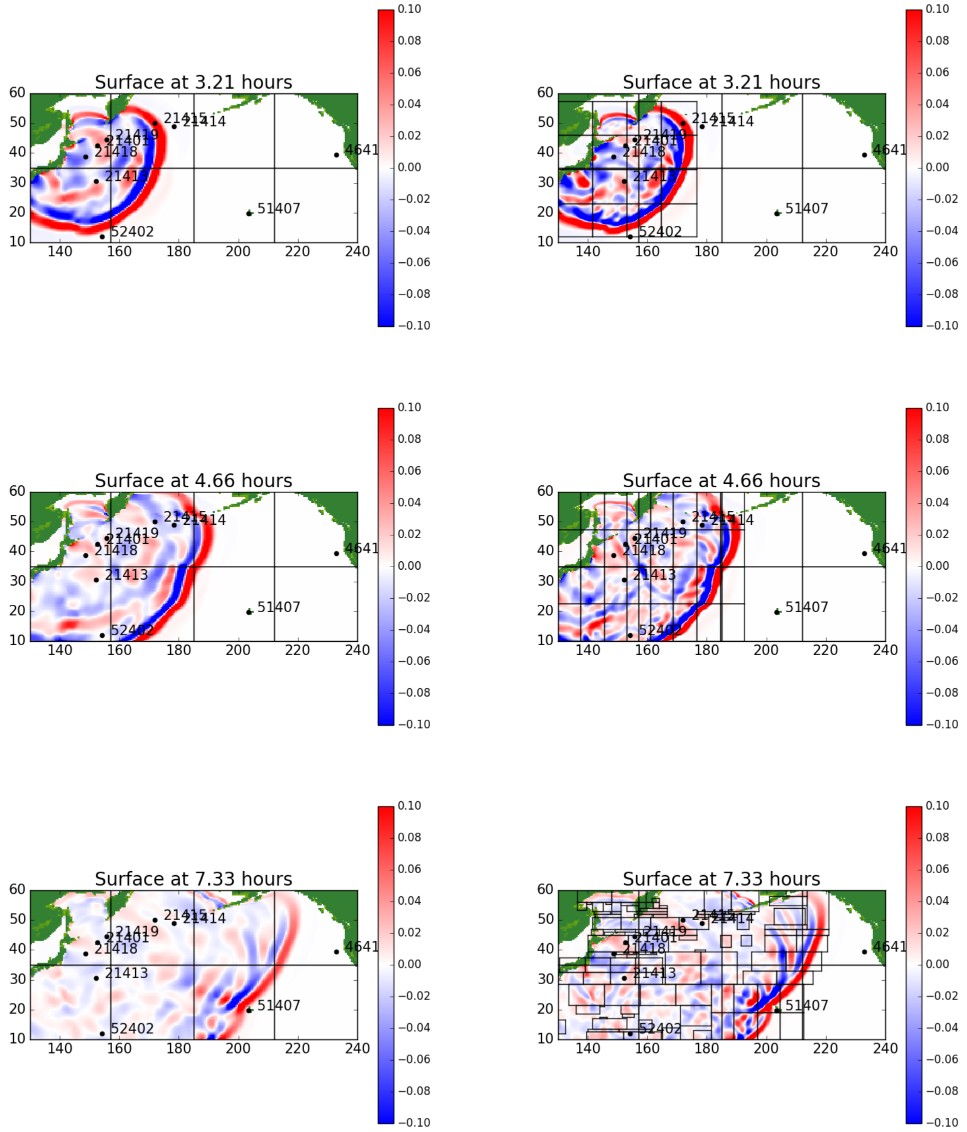


Figure 6.23: Topography changes at the fault region for Tohoku tsunami event

6.3.2 Reference solution

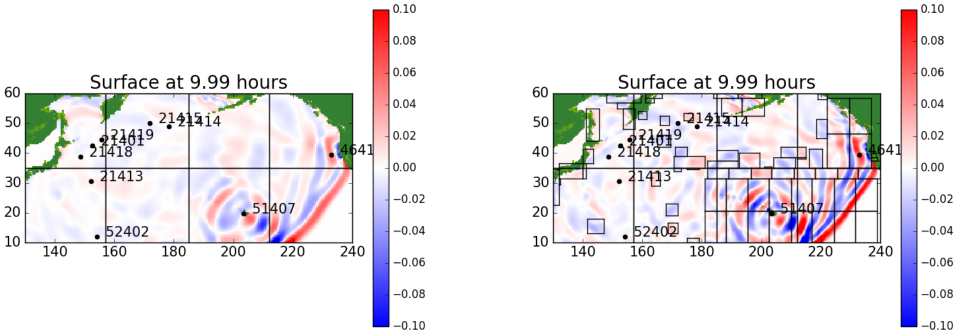
Unlike Chile tsunami test case, the Tohoku earthquake is modeled using multiple subfault slips. With the computed deformation from the best known bathymetry details, the solution for a single level mesh and mesh with maximum AMR level set to two levels are shown in Figure 6.24. The water surface elevation at various gauges for the two level mesh is shown in Figure 6.25. It is evident that the numerical solution is successful in capturing the initial jump in water surface elevation at various gauges. Thereafter, the numerical solution does not follow the gauge measurements closely. See for instance, gauges 21414, 21415 and 21419.



(a) Maximum AMR level 1

(b) Maximum AMR level 2

Figure 6.24: Snapshots of water surface elevation of Tohoku tsunami test case at gauges using the available fault parameters at time 3.21 hours, 4.66 hours and 7.33 hours



(a) Maximum AMR level 1

(b) Maximum AMR level 2

Figure 6.24, cont: Snapshots of water surface elevation of Tohoku tsunami test case at gauges using the available fault parameters at 9.99 hours

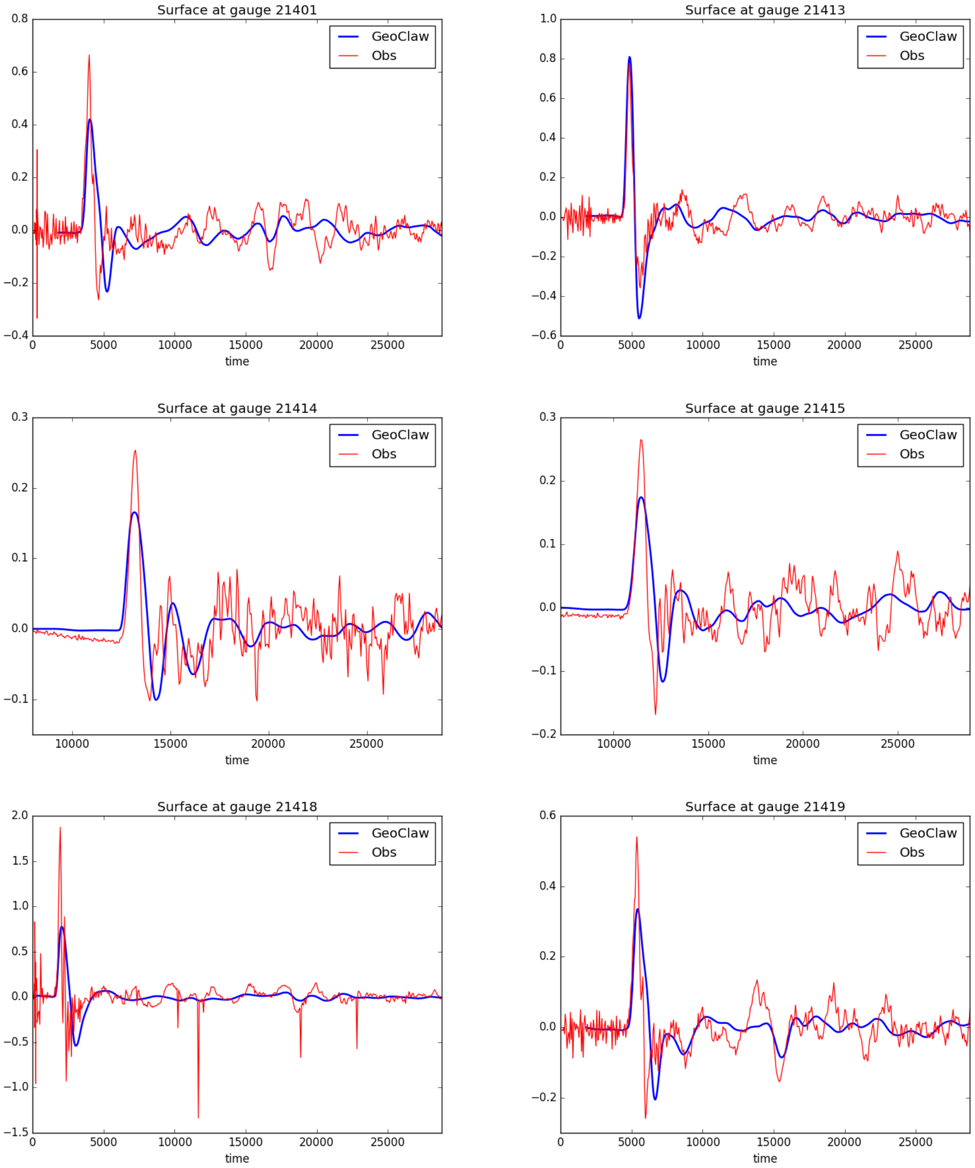


Figure 6.25: Comparison of gauge measurements and numerical results for Tohoku test case using best known bathymetry parameters at gauges 21401, 21413, 21414, 21415, 21418 and 21419

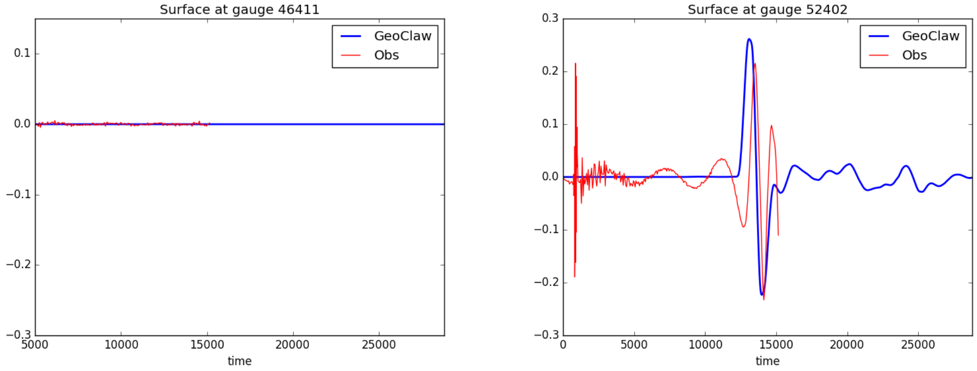


Figure 6.25, cont: Comparison of gauge measurements and numerical results for Tohoku test case using best known bathymetry parameters at gauges 46411 and 52402.

The reference solution in Figure 6.24a indicates that the tsunami that is generated off the coast of Japan and propagates towards the east. The gauge at 51407 is off the Hawaii coast and an important region of interest. For the simulation with maximum AMR set to 1, the simulation looks straightforward. The leading wave traveling in the south-east direction towards Hawaii undergoes a reflection. The reference solution in Figure 6.24b gives a finer solution as the leading wave is traveling and further ensures refinement at the Hawaii region.

6.3.3 Assimilated solution

In the Chile tsunami test case, we used the twin experiment method to calibrate the model and simultaneously demonstrate that assimilation improves the solution. To validate the same, we incorporate the lessons used in the Chile tsunami test case. It is found that a localization radius of $15°$ and a covariance inflation of

0.6 works well for Tohoku test case.

The objective is to assimilate measurement values from certain gauges and observe an improvement in the solution at other gauges. Specifically, the gauge values at $21419, 21418, 21414$ are assimilated, while improvement in solution is to be observed at gauges $21413, 21401, 52402, 21415$. The gauges are chosen such that among the gauges that are most closely located, one gauge value is assimilated, while the other is observed.

Following the ensemble generation procedure, 100 perturbed subfaults are first simulated till the checkpoint time of 2 hours. The ensembles are generated such that from the initial ensemble of 100 states, an eigen decomposition is performed that reveal directions along which the variance among the ensemble members is maximum. Hence, perturbations in these directions are then provided. The initial ensemble of 100 is found to be reasonable as the eigen decomposition of even larger ensemble numbers (upto 500) did not reveal any change in the eigen modes. From the largest eigen modes, 16 ensemble members are created. For illustration, Figure 6.26 shows 9 ensemble members.

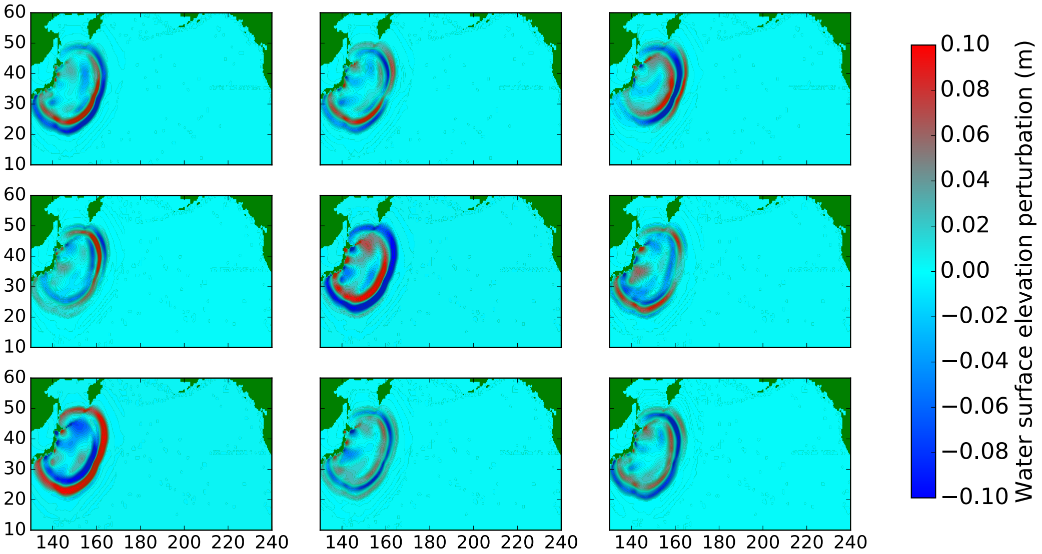
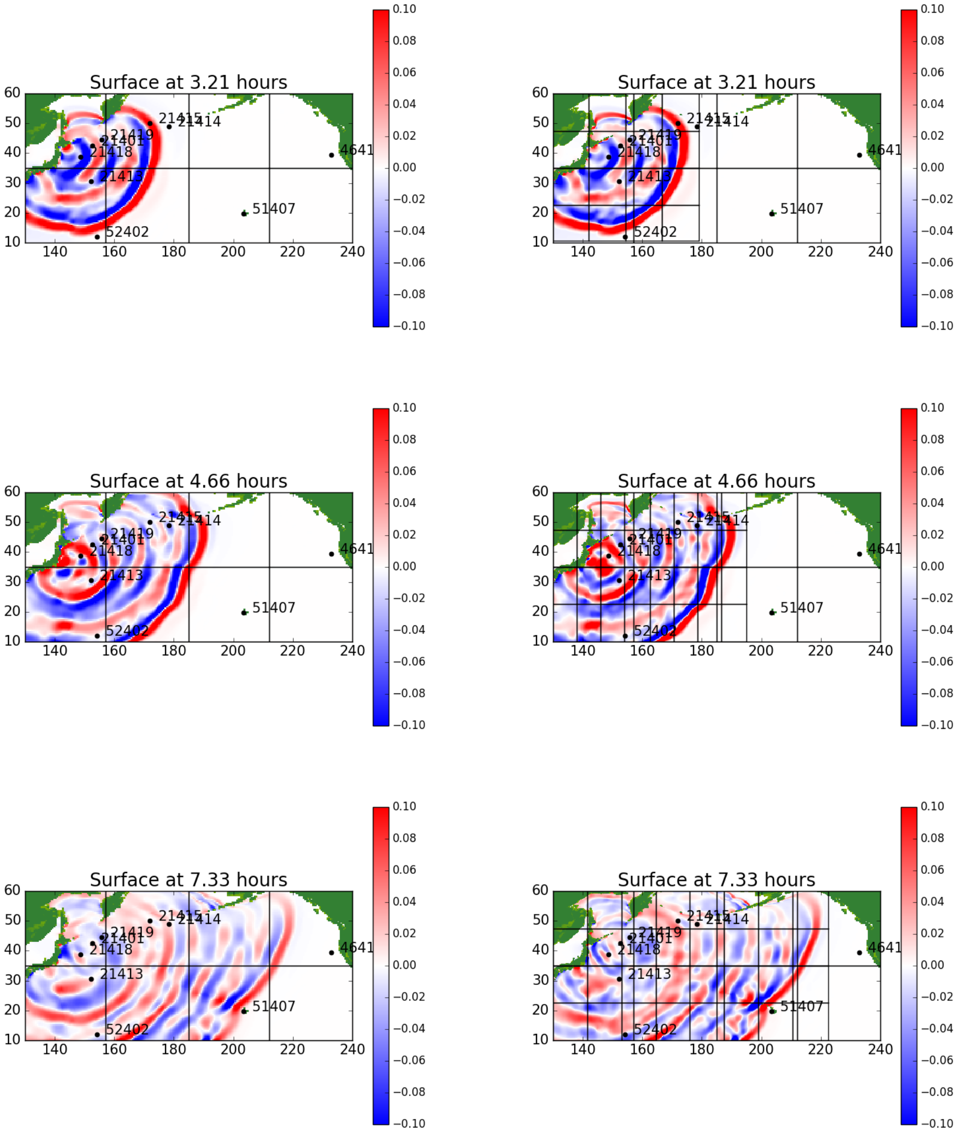


Figure 6.26: Ensemble members for Tohoku test case generated at the checkpoint time

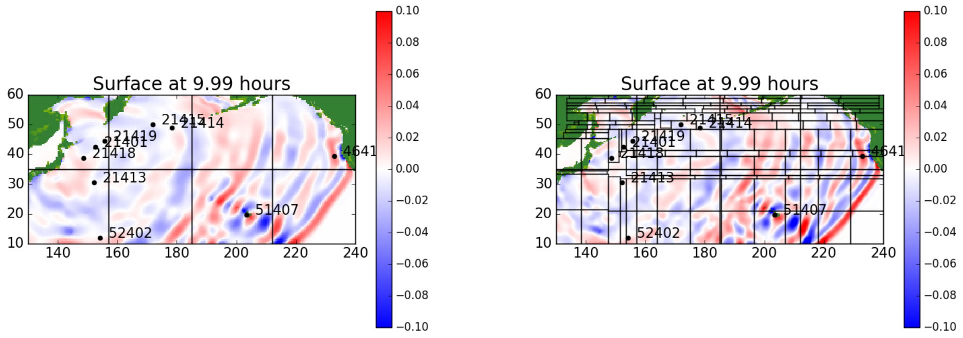
Figure 6.27 show the assimilated solution for level 1 and level 2 meshes. Figure 6.29 shows the corresponding uncertainty in the water surface elevation. For illustration, the simulation with a single level mesh is provided. Similar results are seen for maximum AMR level of 2. Figure 6.28 show the gauge values for the assimilated solution.



(a) Maximum AMR level 1

(b) Maximum AMR level 2

Figure 6.27: Snapshots of assimilated water surface elevation of Tohoku tsunami test case for simulations with maximum AMR set to 1 and 2 at time 3.21 hours, 4.66 hours and 7.33 hours.



(a) Maximum AMR level 1

(b) Maximum AMR level 2

Figure 6.27, cont: Snapshots of assimilated water surface elevation of Tohoku tsunami test case for simulations with maximum AMR set to 1 and 2 at 9.99 hours.

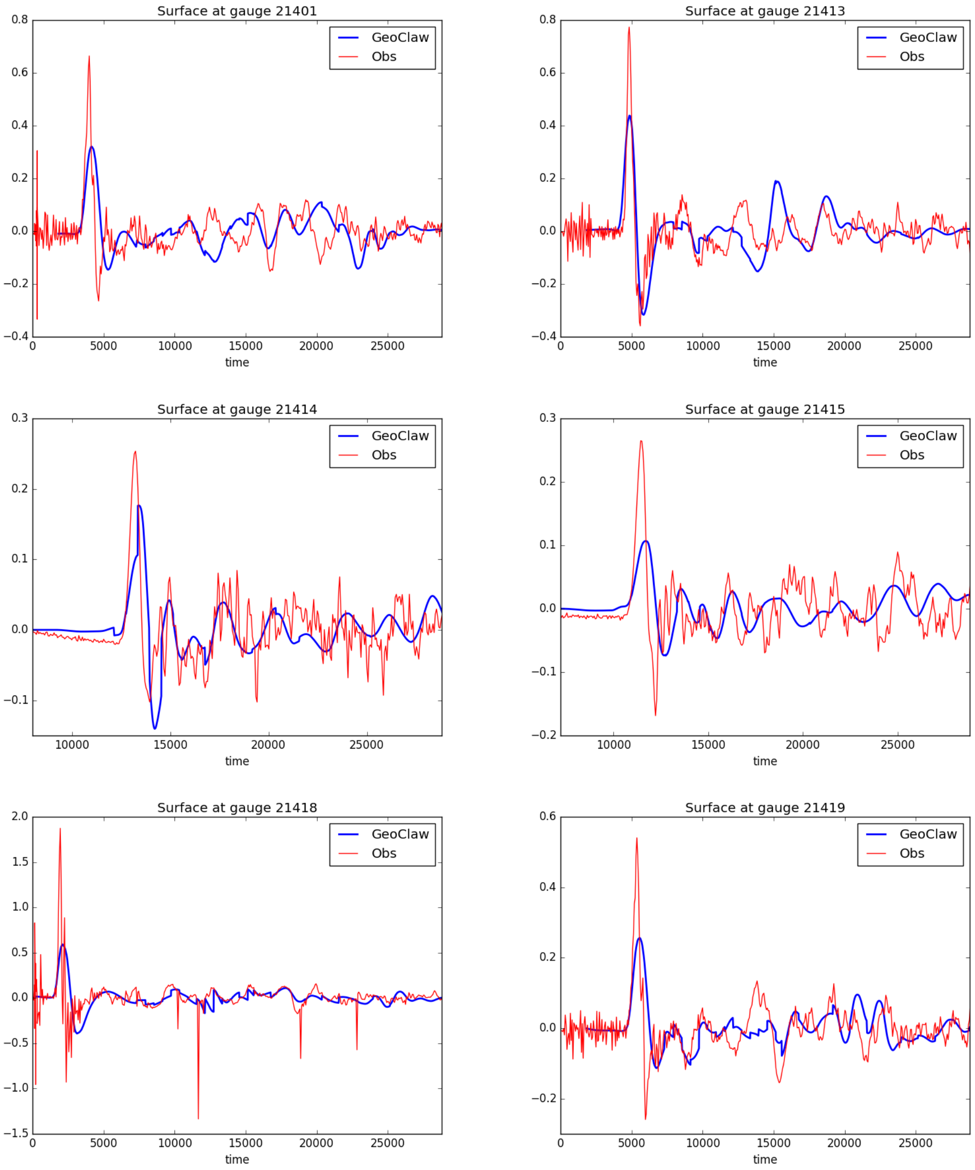


Figure 6.28: Comparison of gauge measurements and assimilated water surface elevation at gauges 21401, 21413, 21414, 21415, 21418 and 21419

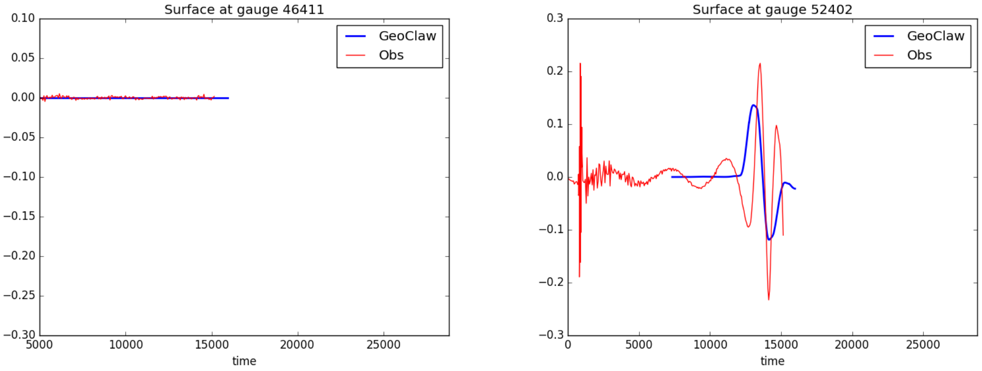


Figure 6.28, cont: Comparison of gauge measurements and assimilated water surface elevation at gauges 46411 and 52402

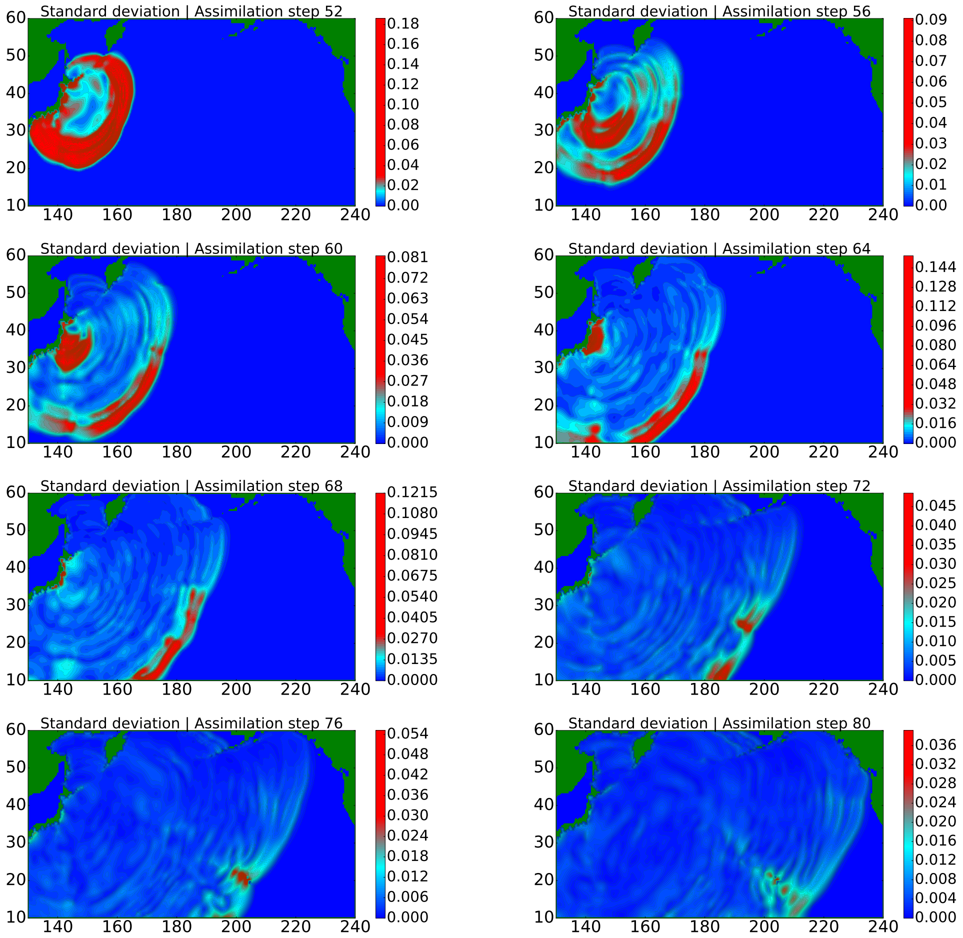


Figure 6.29: Uncertainty in the assimilated water surface elevation for Tohoku test case

The following remarks can be drawn from the assimilated solution:

Remark i: The assimilated solution in Figure 6.27b indicates that the overall level 2 refinement regions are higher than the same simulation without the assimilation. The reason is that the initial ensemble generated from the eigen modes is chosen to have a high variance. By reducing the inflation factor of the initial ensemble member, it is found that the ensemble members have a lower variance and since the members do not differ significantly, the refinement regions are also similar. But it should be noted that an initial ensemble with low variance is prone to collapse as assimilation is repeated over time.

Remark ii: The gauge values in the assimilated solution seem to follow the measurement values more closely than the simulation without assimilation. However, the evidence does not seem to be strong. It must be noted that the initial peak in the water surface elevation at various gauges is not captured as well as the original numerical solution. This is attributed to multiple reasons. One of the reasons is that the plot shows the mean state of the system at the first assimilation step that contains the highest uncertainty. In this case, it happens that the mean water surface elevation after the first assimilation step represents a tsunami of lower amplitude. This is actually not of concern as further assimilation eventually improves the solution, while decreasing the uncertainty in the water surface elevation.

Remark iii: In Figure 6.29, as the leading wave is traveling, the uncertainty in the water surface elevation reduces. Some uncertainty exists near the Hawaii region as expected.

6.3.4 Design of experiment

Various test cases so far indicate that assimilation is effective when the path of the leading wave is along the direction of the observation gauges. The efficacy of the data assimilation in the previous section that assimilates real gauge values seems to be inconclusive. This is possibly due to the placement of the observation gauges mainly being near the coast and lack of gauges placed in the deep ocean.

6.3.4.1 Experiment setup

In a bid to validate the hypothesis regarding placement of gauges along the direction of the leading wave offering improved assimilation, we design a hypothetical observation configuration as shown in Figure 6.30. We assimilate gauge 2 and gauge 3 and observe the solution at the remaining gauges.

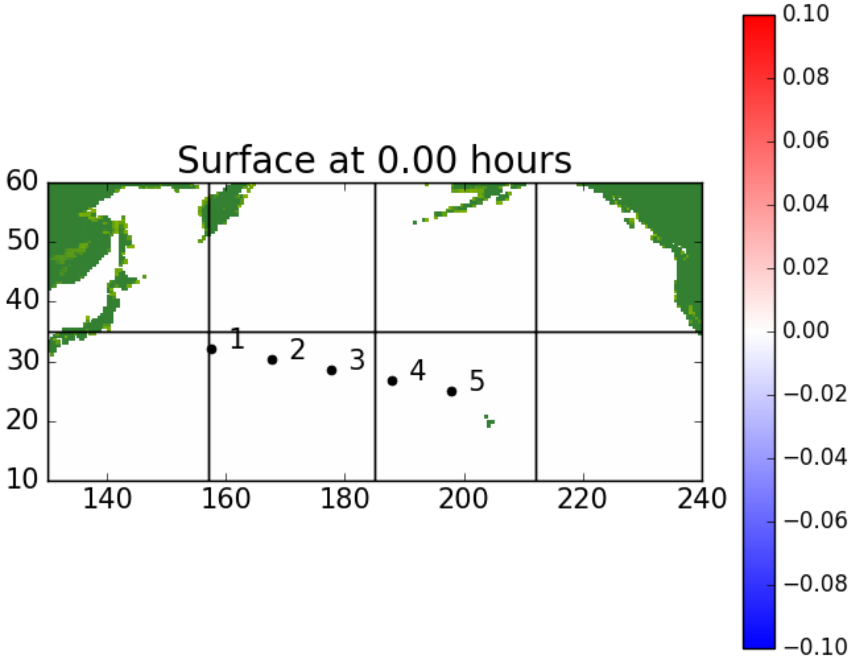


Figure 6.30: Hypothetical observation configuration

6.3.4.2 True solution and free run

Similar to the previous experiments, the “true" observation values will be simulated and will be assimilated into the simulation with a different earthquake source. The “true" gauge values are generated from a tsunami with the earthquake source given by [4]. On the other hand, we take another earthquake source given by `http://www.geol.ucsb.edu/faculty/ji/big_earthquakes/2011/03/0311_v3/Honshu.html`. It should be noted that there are usually various earthquake source models. For the current design, instead of perturbing the Okada parameters, we have adopted simulations of various earthquake source models. The tsunami event for the truth and free run is shown in Figure 6.31. The figure show that the true tsunami event produced a tsunami with a higher amplitude of the leading wave than the free run.

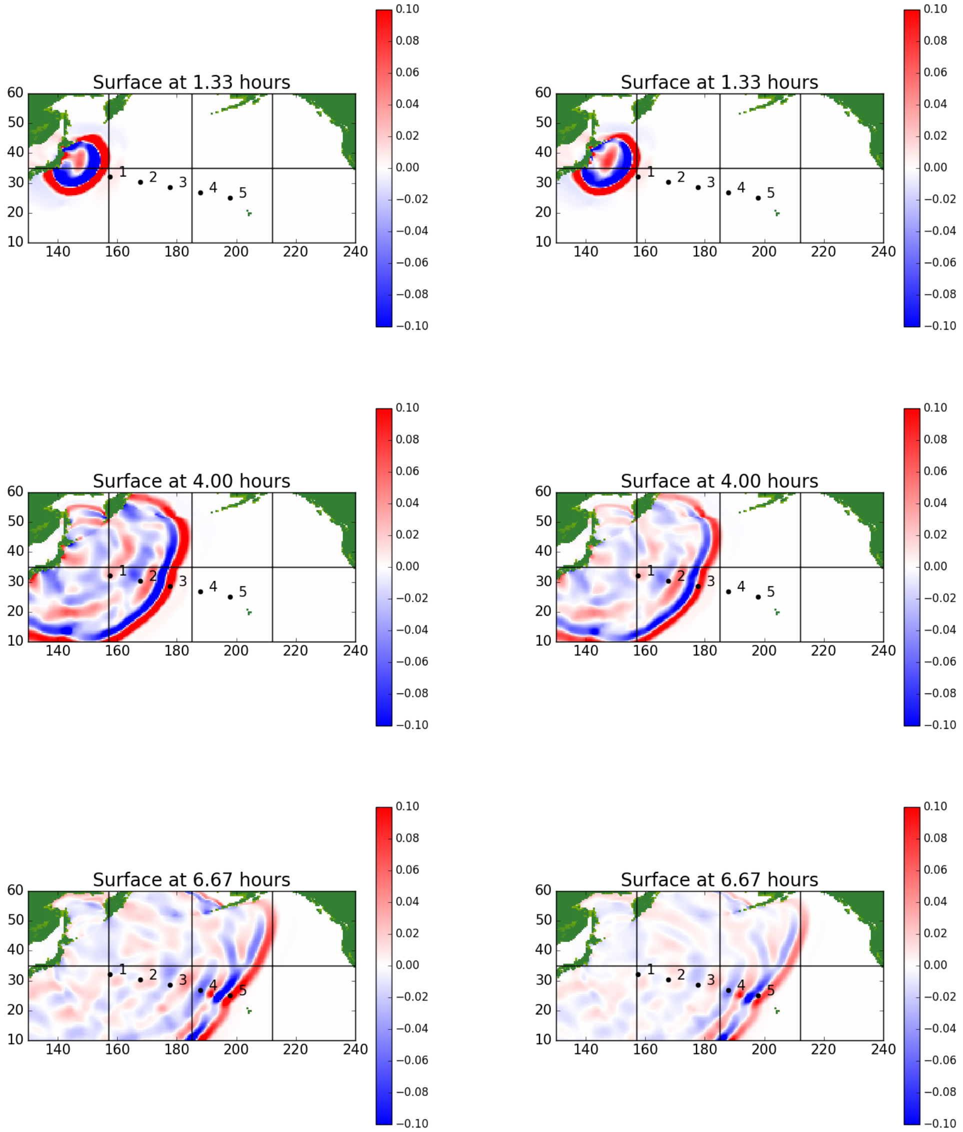


Figure 6.31: Water surface elevation comparison of (Right) free run state and (Left) true state for experimental observation configuration setup at time 1.33 hours, 4.00 hours and 6.67 hours

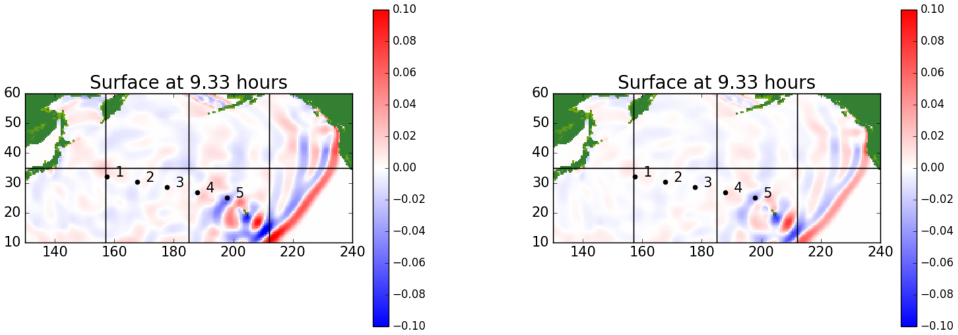


Figure 6.31, cont: Water surface elevation comparison of (Right) free run state and (Left) true state for experimental observation configuration setup at time 9.33 hours

6.3.4.3 Assimilated solution

Assimilation is performed using the LETKF filter with a localization radius of 20° and inflation factor of 0.9 and observation root mean square error of 0.005 meters. Figure 6.32 demonstrates the assimilated solution, wherein the same free run that had earlier produced a weaker tsunami than truth (see amplitude of leading wave), is now comparable to the truth after assimilation. Further, the gauge plots in Figure 6.33 validate that assimilation of solution at gauge 2 and gauge 3 definitely improves the solution at gauges 4 and 5. Gauge 1 solution is worse than the free run. This is expected because, the leading wave first passes through gauge 1 and then starts to assimilate from gauge 2 onwards. Moreover, the leading wave near the coast typically has higher uncertainty in the water surface elevation.

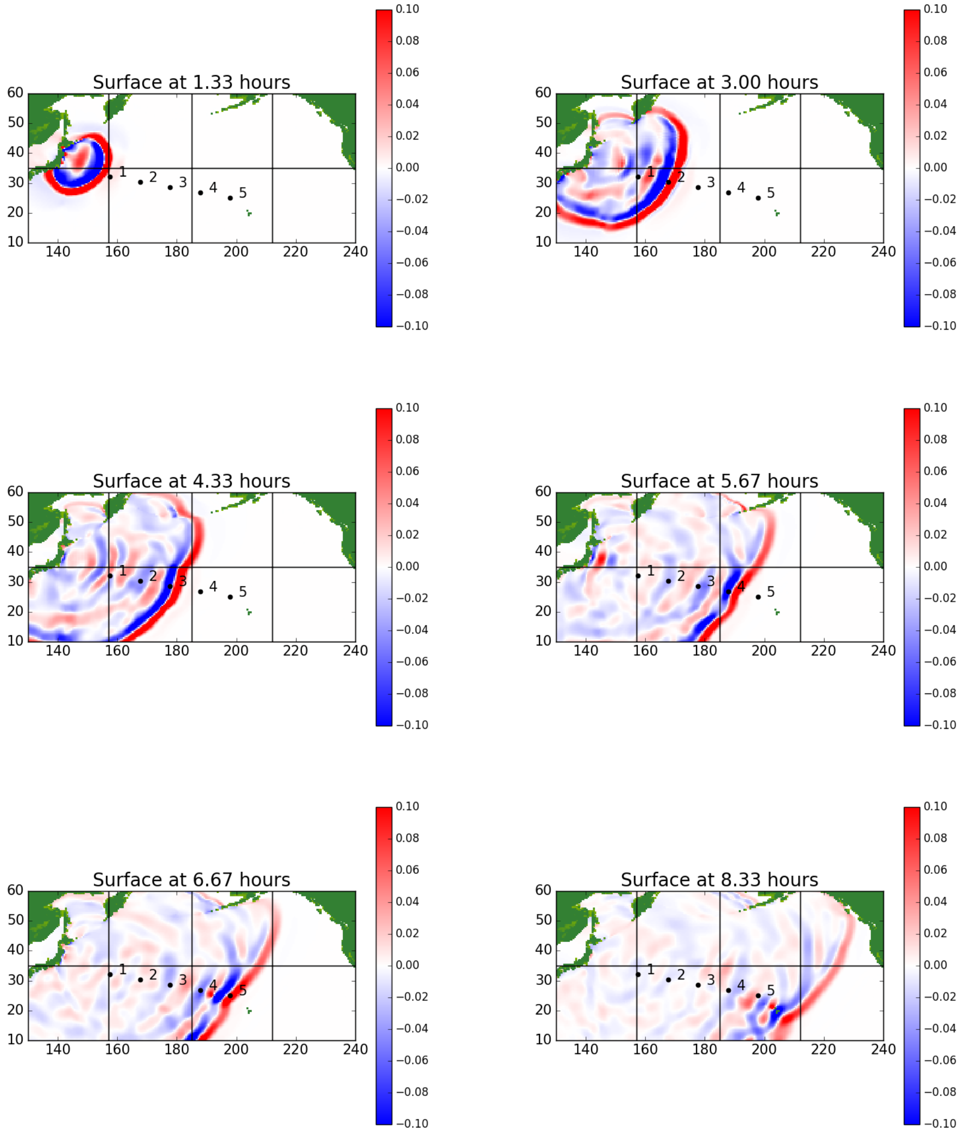


Figure 6.32: Water surface elevation of assimilated solution in the experimental observation configuration setup at time 1.33 hours, 3.00 hours, 4.33 hours, 5.67 hours, 6.67 hours and 8.33 hours

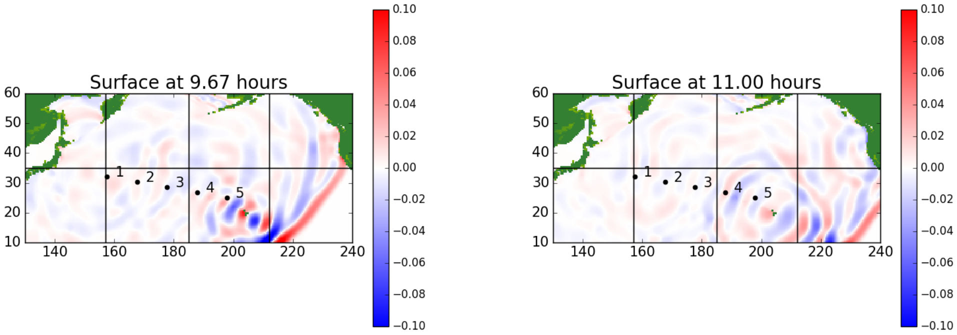


Figure 6.32, cont: Water surface elevation of assimilated solution in the experimental observation configuration setup at time 9.67 hours and 11.00 hours

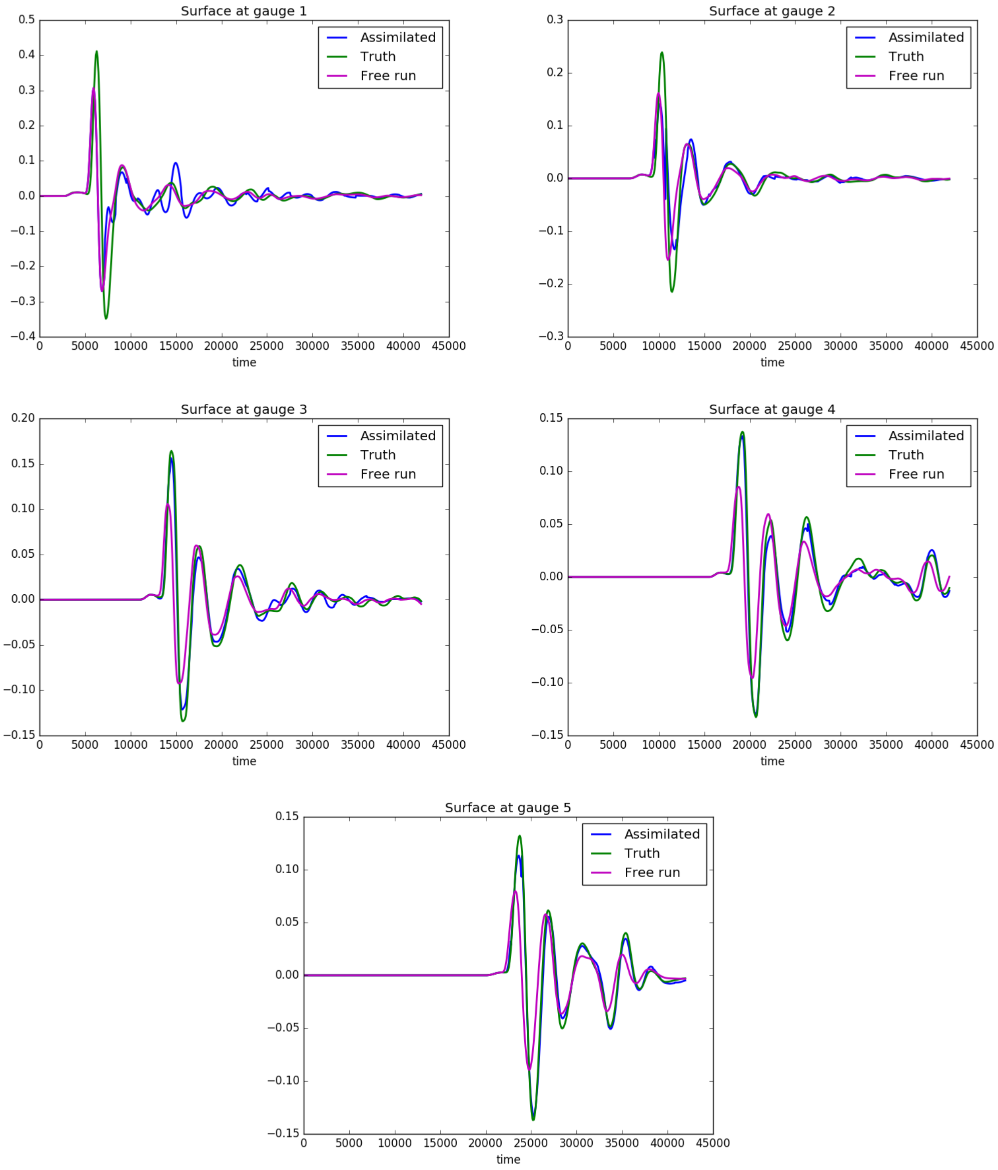


Figure 6.33: Comparison of gauge measurements for truth, free run and assimilated water surface elevation for the experimental observation configuration setup

Chapter 7

Summary

We have systematically developed a data assimilation system by coupling two components - the numerical model, GeoClaw and the ensemble based data assimilation filters using the library, PDAF. Through the literature review, we discussed the historical and contemporary trends in using data assimilation techniques with numerical models for large scale geophysical flows. Subsequently, data assimilation techniques in particular are discussed in detail. The development of Kalman filters and the motivation to pursue ensemble based Kalman filters, specifically the square root filters is understood systematically.

The other component of the data assimilation system - the numerical model, GeoClaw is explained. The unique characteristic of GeoClaw is the use of multi-level adaptive mesh refinement technique, that offers the capability to obtain refined solutions at the regions of interest and dynamically adapts with the flow dynamics. The adaptive mesh algorithm is discussed, that lays the foundation for the discussion on coupling it with the ensemble based filters.

The coupling of both the components is the highlight of the research. The sequential process of forecasting and the update runs into a problem when the numerical model uses adaptive mesh refinement. The reason being that every ensemble member will adapt in its way with time, thereby making it impossible to compute quantities such as the mean state and the variance of the state at the update step. This situation is tackled using the supermesh technique. The technique projects all the ensemble members on a common mesh and then updates the solution. A regridding operation before the next forecast ensures that the individual ensemble members are coarsened in areas where the refinement is unnecessary. Thus, as the assimilation system runs, over-refined ensemble members are avoided.

In the process of implementing the technique for tsunami simulation applications, it was found that the solutions were blowing up at assimilations steps. This was attributed to the methodology of generating initial ensemble members using the perturbations in the slip in the bathymetry. This was countered by a suitable ensemble generation methodology. The method calculates the eigen modes of the covariance matrix of states of a large ensemble size, simulated up to the checkpoint time . The eigen modes indicate the 'directions' of maximum variance. Perturbation are then provided in these directions and new ensembles are thus created.

We performed various twin experiments assessing the effect of observation configuration with respect to number of observations, localization radius and ensemble size on errors between assimilated state and free run. This indicates that assimilation is effective. Additionally, we showed that more accurate state estimates can be obtained while reducing uncertainty with assimilation steps. This is demonstrated through three test cases - *Radial bowl test case*, *Chile tsunami test case* and *Tohoku tsunami test case*.

During the forecasting of large scale geophysical flows, there exists errors such as the model error, discretization error, error in initial conditions, error in boundary condition etc. These errors are unavoidable and induced due to the underlying assumptions of the model and in the process of translating the governing equations of fluid flow to computer codes for simulations. The research demonstrates these errors can be effectively mitigated by use of data assimilation. The *Radial bowl test case* analyzes the efficacy of data assimilation on reducing the discretization error and the initial condition error. The *Chile tsunami test case* analyzes various aspects of ensemble based filters such as localization radius, number of ensembles required, observation configuration etc. Further, an algorithm for generating ensemble members for tsunami model is discussed. The method maps the uncertainty in the bathymetry changes due to the associated earthquake, to the uncertainty in the initial condition of the event. The lessons from *Chile tsunami test case* are imbibed in *Tohoku tsunami test case*, where multiple gauge measurements are available. This serves as a validation case, wherein the efficacy of data assimilation is proved by assimilating some gauge measurements, while observing the improvement in solution at the other gauges.

In both the tsunami test cases, we concurrently showed the execution of assimilation with adaptive mesh refinement up to 2 levels. We demonstrated that the AMR capability captured the oscillations in the assimilated state and tracked the waves as they traveled towards the coast. Further, we demonstrated a hypothetical observation configuration for improved assimilation. This was reaffirmed with the estimated water surface elevation closely following the true values at the simulated gauges. The AMR-coupled data assimilation framework has been tested with higher refinement levels (of up to 5). The results presented here are generated from the code for GEOCLAW at `https://zenodo.org/badge/latestdoi/30735579` and the code for AMRCLAW at `https://zenodo.org/badge/latestdoi/92109632`.

Chapter 8

Conclusion

The study is aimed to provide some primary insights in the development of data driven computational framework using AMR capable numerical models. In particular, ensemble data assimilation requires integration of several model runs, This can be computationally expensive for large scale hydrodynamic models. On the contrary, the numerical codes with AMR technique offer advantages with respect to computational speed as areas that are not of interest such as the deep ocean can retain coarse mesh. Hence, ensemble data assimilation with AMR is ideal for use in real-time forecasting scenarios, where time of simulations is of essence and may serve as a bottleneck for initiating evacuation measures. The coupling methodology demonstrated in this study, can be extended to various models currently in use. The possible areas to further explore are storm surges and wave runup. This will aid in the development of a potential real-time forecasting tool, that offers computational speed and uncertainty quantification simultaneously.

The implementation of adaptive data-driven simulation of tsunamis is published in [55].

Bibliography

[1] Brent W. Ache, Kristen M. Crossett, Percy A. Pacheco, Jeffery E. Adkins, and Peter C. Wiley. "The coast" is complicated: A model to consistently describe the nation's coastal population. *Estuaries and coasts*, 38(1):151–155, 2015.

[2] Altuğ Aksoy, David C. Dowell, and Chris Snyder. A multicase comparative assessment of the ensemble Kalman filter for assimilation of radar observations. Part I: Storm-scale analyses. *Monthly Weather Review*, 137(6):1805–1824, 2009.

[3] J. I. Allen, P. J. Somerfield, and F. J. Gilbert. Quantifying uncertainty in high-resolution coupled hydrodynamic-ecosystem models. *Journal of Marine Systems*, 64(1–4):3 – 14, 2007.

[4] Charles J Ammon, Thorne Lay, Hiroo Kanamori, and Michael Cleveland. A rupture model of the 2011 off the pacific coast of tohoku earthquake. *Earth, Planets and Space*, 63(7):33, 2011.

[5] J. L. Anderson. Ensemble Kalman filters for large geophysical applications. *IEEE Control Systems*, 29(3):66–82, June 2009.

[6] Jeffrey Anderson, Tim Hoar, Kevin Raeder, Hui Liu, Nancy Collins, Ryan Torn, and Avelino Avellano. The Data assimilation Research Testbed: A

community facility. *Bulletin of the American Meteorological Society*, 90(9):1283–1296, 2009.

[7] Jeffrey L. Anderson. An Ensemble Adjustment Kalman Filter for Data Assimilation. *Monthly Weather Review*, 129(12):2884–2903, 2001.

[8] Jeffrey L. Anderson. A local least squares framework for ensemble filtering. *Monthly Weather Review*, 131(4):634–642, 2003.

[9] Jeffrey L. Anderson. An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A*, 59(2):210–224, 2007.

[10] Jeffrey L Anderson and Stephen L Anderson. A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12):2741–2758, 1999.

[11] Ivo Babushka, Jagdish Chandra, and Joseph E. Flaherty. *Adaptive computational methods for partial differential equations*, volume 16. Siam, 1983.

[12] Deborah Balk, Mark R Montgomery, Gordon McGranahan, Donghwan Kim, Valentina Mara, Megan Todd, Thomas Buettner, and Audrey Dorélien. Mapping urban settlements and the risks of climate change in Africa, Asia and South America. *Population dynamics and climate change*, 80, 2009.

[13] Randolph E. Bank, Andrew H. Sherman, and Alan Weiser. Some refinement algorithms and data structures for regular local mesh refinement. *Sci. Comput. Appl. Math. Comput. Phys. Sci*, 1:3–17, 1983.

[14] R. N. Bannister. A review of forecast error covariance statistics in atmospheric variational data assimilation. I: Characteristics and measurements of forecast error covariances. *Quarterly Journal of the Royal Meteorological Society*, 134(637):1951–1970, 2008.

[15] M. Berger and I. Rigoutsos. An algorithm for point clustering and grid generation. *IEEE Transactions on Systems, Man, and Cybernetics*, 21(5):1278–1286, Sep 1991.

[16] M. J. Berger, D. L. George, R. J. LeVeque, and K. T. Mandli. The GeoClaw software for depth-averaged flows with adaptive refinement. *Advances in Water Resources*, 34:1195–1206, September 2011.

[17] Marsha J. Berger and Joseph Oliger. Adaptive mesh refinement for hyperbolic partial differential equations. *Journal of Computational Physics*, 53(3):484 – 512, 1984.

[18] Craig H. Bishop, Brian J. Etherton, and Sharanya J. Majumdar. Adaptive sampling with the Ensemble Transform Kalman Filter. Part I: Theoretical aspects. *Monthly Weather Review*, 129(3):420–436, 2001.

[19] Rainer Bleck. An oceanic general circulation model framed in hybrid isopycnic-cartesian coordinates. *Ocean Modelling*, 4(1):55 – 88, 2002.

[20] James D. Brown, Tom Spencer, and Iris Moeller. Modeling storm surge flooding of an urban area with particular reference to modeling uncertainties:

A case study of Canvey Island, United Kingdom. *Water Resources Research*, 43(6), 2007.

[21] Gerrit Burgers, Peter Jan van Leeuwen, and Geir Evensen. Analysis scheme in the ensemble kalman filter. *Monthly weather review*, 126(6):1719–1724, 1998.

[22] Carsten Burstedde, Lucas C. Wilcox, and Omar Ghattas. p4est: Scalable algorithms for parallel adaptive mesh refinement on forests of octrees. *SIAM Journal on Scientific Computing*, 33(3):1103–1133, 2011.

[23] T. Butler, M. U. Altaf, C. Dawson, I. Hoteit, X. Luo, and T. Mayo. Data Assimilation within the Advanced Circulation (ADCIRC) Modeling Framework for Hurricane Storm Surge Forecasting. *Monthly Weather Review*, 140(7):2215–2231, 2012.

[24] V. Carmillet, J.-M. Brankart, P. Brasseur, H. Drange, G. Evensen, and J. Verron. A singular evolutive extended Kalman filter to assimilate ocean color data in a coupled physical–biochemical model of the North Atlantic ocean. *Ocean Modelling*, 3(3):167 – 192, 2001.

[25] Eric P. Chassignet, Harley E. Hurlburt, Ole Martin Smedstad, George R. Halliwell, Patrick J. Hogan, Alan J. Wallcraft, Remy Baraille, and Rainer Bleck. The {HYCOM} (HYbrid Coordinate Ocean Model) data assimilative system. *Journal of Marine Systems*, 65(1–4):60 – 83, 2007.

[26] P. Courtier, E. Andersson, W. Heckley, D. Vasiljevic, M. Hamrud, A. Hollingsworth, F. Rabier, M. Fisher, and J. Pailleux. The ECMWF implementation of three-dimensional variational assimilation (3D-Var). I: Formulation. *Quarterly Journal of the Royal Meteorological Society*, 124(550):1783–1807, 1998.

[27] P. Courtier, J.-N. Thépaut, and A. Hollingsworth. A strategy for operational implementation of 4D-Var, using an incremental approach. *Quarterly Journal of the Royal Meteorological Society*, 120(519):1367–1387, 1994.

[28] Ralf Deiterding. *Detonation Structure Simulation with AMROC*, pages 916–927. Springer Berlin Heidelberg, Berlin, Heidelberg, 2005.

[29] Donald W. Denbo, K. T. McHugh, J. R. Osborne, P. Sorvik, and A. J. Venturato. NOAA tsunami forecasting system: Design and implementation using service oriented architecture. In *23rd Conference on Interactive Information Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, 2007.

[30] Juan Du, Jiang Zhu, Fangxin Fang, C. C. Pain, and I. M. Navon. Ensemble data assimilation applied to an adaptive mesh ocean model. *International Journal for Numerical Methods in Fluids*, 82(12):997–1009, 2016.

[31] Christopher A. Edwards, Andrew M. Moore, Ibrahim Hoteit, and Bruce D. Cornuelle. Regional ocean data assimilation. *Annual review of marine science*, 7:21–42, 2015.

[32] Martin Ehrendorfer. A review of issues in ensemble-based Kalman filtering. *Meteorologische Zeitschrift*, 16(6):795–818, 12 2007.

[33] Geir Evensen. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, 99(C5):10143–10162, 1994.

[34] Geir Evensen. The Ensemble Kalman Filter: :Theoretical formulation and practical implementation. *Ocean Dynamics*, 53(4):343–367, 2003.

[35] Geir Evensen and Peter Jan van Leeuwen. Assimilation of Geosat altimeter data for the Agulhas current using the Ensemble Kalman filter with a quasi-geostrophic model. *Monthly Weather Review*, 124(1):85–96, 1996.

[36] F. Fang, C. C. Pain, I. M. Navon, M. D. Piggott, G. J. Gorman, P. E. Farrell, P. A. Allison, and A. J. H. Goddard. A POD reduced-order 4D-Var adaptive mesh ocean modelling approach. *International Journal for Numerical Methods in Fluids*, 60(7):709–732, 2009.

[37] Joseph E. Flaherty and Peter K. Moore. Integrated space-time adaptive hp-refinement methods for parabolic systems. *Applied Numerical Mathematics*, 16(3):317 – 341, 1995.

[38] Gregory Gaspari and Stephen E Cohn. Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society*, 125(554):723–757, 1999.

[39] David L. George. Augmented Riemann solvers for the shallow water equations over variable topography with steady states and inundation. *Journal of Computational Physics*, 227(6):3089 – 3113, 2008.

[40] David L. George and Randall J. LeVeque. Finite volume methods and adaptive refinement for global tsunami propagation and local inundation. 2006.

[41] Georg A. Grell and Dezső Dévényi. A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. *Geophysical Research Letters*, 29(14):38–1–38–4, 2002.

[42] Aditya Riadi Gusman, Anne F. Sheehan, Kenji Satake, Mohammad Heidarzadeh, Iyan Eka Mulia, and Takuto Maeda. Tsunami data assimilation of Cascadia seafloor pressure gauge records from the 2012 Haida Gwaii earthquake. *Geophysical Research Letters*, 43(9):4189–4196, 2016.

[43] Thomas M. Hamill. Ensemble-based atmospheric data assimilation. *Predictability of weather and climate*, pages 124–156, 2006.

[44] Thomas M Hamill, Jeffrey S Whitaker, and Chris Snyder. Distance-dependent filtering of background error covariance estimates in an ensemble kalman filter. *Monthly Weather Review*, 129(11):2776–2790, 2001.

[45] A. W. Heemink, M. Verlaan, and A. J. Segers. Variance reduced Ensemble Kalman filtering. *Monthly Weather Review*, 129(7):1718–1728, 2001.

[46] Jason T. Holt, J. Icarus Allen, Roger Proctor, and Francis Gilbert. Error quantification of a high-resolution coupled hydrodynamic–ecosystem coastal–ocean

model: Part 1 model overview and assessment of the hydrodynamics . *Journal of Marine Systems*, 57(1–2):167 – 188, 2005.

[47] I. Hoteit, D.-T. Pham, M. E. Gharamti, and X. Luo. Mitigating observation perturbation sampling errors in the stochastic EnKF. *Monthly Weather Review*, 143(7):2918–2936, 2015.

[48] I. Hoteit, G. Triantafyllou, and G. Korres. Using low-rank Ensemble Kalman filters for data assimilation with high dimensional imperfect models. *JNAIAM*, 2(1-2):67–78, 2007.

[49] Ibrahim Hoteit, Tim Hoar, Ganesh Gopalakrishnan, Nancy Collins, Jeffrey Anderson, Bruce Cornuelle, Armin Köhl, and Patrick Heimbach. A MITgcm/DART ensemble analysis and prediction system with application to the Gulf of Mexico. *Dynamics of Atmospheres and Oceans*, 63:1 – 23, 2013.

[50] Ibrahim Hoteit, George Korres, and George Triantafyllou. Comparison of extended and ensemble based Kalman filters with low and high resolution primitive equation ocean models. *Nonlinear Processes in Geophysics*, 12(5):755–765, 2005.

[51] Ibrahim Hoteit, Dinh-Tuan Pham, and Jacques Blum. A simplified reduced order Kalman filtering and application to altimetric data assimilation in Tropical Pacific. *Journal of Marine Systems*, 36(1–2):101 – 127, 2002.

[52] P. L. Houtekamer and Herschel L. Mitchell. A sequential Ensemble Kalman

filter for atmospheric data assimilation. *Monthly Weather Review*, 129(1):123–137, 2001.

[53] P. L. Houtekamer and Fuqing Zhang. Review of the Ensemble Kalman filter for atmospheric data assimilation. *Monthly Weather Review*, 144(12):4489–4532, 2016.

[54] Brian R. Hunt, Eric J. Kostelich, and Istvan Szunyogh. Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D: Nonlinear Phenomena*, 230(1–2):112 – 126, 2007.

[55] Pushkar Kumar Jain, Kyle Mandli, Ibrahim Hoteit, Omar Knio, and Clint Dawson. Dynamically adaptive data-driven simulation of extreme hydrological flows. *Ocean Modelling*, 122:85 – 103, 2018.

[56] Hrvoje Jasak, Aleksandar Jemcov, Zeljko Tukovic, et al. OpenFOAM: A C++ library for complex physics simulations. In *International workshop on coupled methods in numerical dynamics*, volume 1000, pages 1–20. IUC Dubrovnik, Croatia, 2007.

[57] Andrew H. Jazwinski. Mathematics in Science and Engineering. *Stochastic Processes and Filtering Theory*, 64, 1970.

[58] Rudolph E. Kalman and Richard S. Bucy. New results in linear filtering and prediction theory. *Journal of basic engineering*, 83(3):95–108, 1961.

[59] Rudolph Emil Kalman. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82(1):35–45, 1960.

[60] Christian L. Keppenne. Data assimilation into a primitive-equation model with a parallel Ensemble Kalman filter. *Monthly Weather Review*, 128(6):1971–1981, 2000.

[61] Christian L. Keppenne and Michele M. Rienecker. Initial testing of a massively parallel Ensemble Kalman filter with the Poseidon Isopycnal Ocean General Circulation Model. *Monthly Weather Review*, 130(12):2951–2965, 2002.

[62] Randall J. Leveque. Clawpack: A software package for solving multi–dimensional conservation laws. In *Proc. 5th Intl. Conf. Hyperbolic Problems*, pages 188–197. Citeseer, 1994.

[63] Randall J. LeVeque. *Finite volume methods for hyperbolic problems*, volume 31. Cambridge university press, 2002.

[64] Randall J. LeVeque and David L. George. High-resolution finite volume methods for the shallow water equations with bathymetry and dry states. *Advanced numerical models for simulating tsunami waves and runup*, 10:43–73, 2008.

[65] Zhijin Li, James C. McWilliams, Kayo Ide, and John D. Farrara. A multiscale variational data assimilation scheme: Formulation and illustration. *Monthly Weather Review*, 143(9):3804–3822, 2015.

[66] Knut Arild Lisæter, Julia Rosanova, and Geir Evensen. Assimilation of ice concentration in a coupled ice–ocean model, using the Ensemble Kalman

filter. *Ocean Dynamics*, 53(4):368–388, 2003.

[67] Andrew C. Lorenc. The potential of the Ensemble Kalman filter for NWP—a comparison with 4D-Var. *Quarterly Journal of the Royal Meteorological Society*, 129(595):3183–3203, 2003.

[68] Richard A. Luettich Jr., Johannes J. Westerink, and Norman W. Scheffner. ADCIRC: An advanced three-dimensional circulation model for shelves, coasts, and estuaries. Report 1. Theory and methodology of ADCIRC-2DDI and ADCIRC-3DL. Technical report, DTIC Document, 1992.

[69] Xiaodong Luo and Ibrahim Hoteit. Robust ensemble filtering and its relation to covariance inflation in the Ensemble Kalman filter. *Monthly Weather Review*, 139(12):3938–3953, 2011.

[70] G. Madec. *NEMO ocean engine*. Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No 27, ISSN No 1288-1619, 2008.

[71] Jan Mandel, Loren Cobb, and Jonathan D. Beezley. On the convergence of the Ensemble Kalman filter. *Applications of Mathematics*, 56(6):533–541, Dec 2011.

[72] Kyle T. Mandli, Aron J. Ahmadia, Marsha Berger, Donna Calhoun, David L. George, Yiannis Hadjimichael, David I. Ketcheson, Grady I. Lemoine, and Randall J. LeVeque. Clawpack: building an open source ecosystem for solving hyperbolic pdes. *PeerJ Computer Science*, 2:e68, August 2016.

[73] Talea Mayo, Troy Butler, Clint Dawson, and Ibrahim Hoteit. Data assimilation within the Advanced Circulation (ADCIRC) modeling framework for the estimation of Manning's friction coefficient . *Ocean Modelling*, 76:43 – 58, 2014.

[74] Christian Meinig, Scott E Stalin, Alex I Nakamura, and Hugh B Milburn. Real-time deep-ocean tsunami measuring, monitoring, and reporting system: The noaa dart ii description and disclosure. *NOAA, Pacific Marine Environmental Laboratory (PMEL)*, pages 1–15, 2005.

[75] Lewis Mitchell and Alberto Carrassi. Accounting for model error due to unresolved scales within ensemble kalman filtering. *Quarterly Journal of the Royal Meteorological Society*, 141(689):1417–1428, 2015.

[76] Kazuyuki Nakamura, Tomoyuki Higuchi, and Naoki Hirose. Sequential data assimilation: Information fusion of a numerical simulation and large scale observation data. *J. UCS*, 12(6):608–626, 2006.

[77] Kazuyuki Nakamura, Naoki Hirose, Byung Ho Choi, and Tomoyuki Higuchi. *Particle Filtering in Data Assimilation and Its Application to Estimation of Boundary Condition of Tsunami Simulation Model*, pages 353–366. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[78] L.-J. Natvik and Geir Evensen. Assimilation of ocean colour data into a biochemical model of the North Atlantic: Part 1. Data assimilation experiments. *Journal of Marine Systems*, 40:127–153, 2003.

[79] Lars Nerger and Wolfgang Hiller. Software for ensemble-based data assimilation systems—Implementation strategies and scalability. *Computers & Geosciences*, 55:110 – 118, 2013.

[80] LARS NERGER, WOLFGANG HILLER, and JENS SCHRÖTER. A comparison of error subspace Kalman filters. *Tellus A*, 57(5):715–735, 2005.

[81] Lars Nerger, Tijana Janjić, Jens Schröter, and Wolfgang Hiller. A regulated localization scheme for ensemble-based Kalman filters. *Quarterly Journal of the Royal Meteorological Society*, 138(664):802–812, 2012.

[82] Lars Nerger, Tijana Janjić, Jens Schröter, and Wolfgang Hiller. A unification of Ensemble square root Kalman filters. *Monthly Weather Review*, 140(7):2335–2345, 2012.

[83] Barbara Neumann, Athanasios T. Vafeidis, Juliane Zimmermann, and Robert J. Nicholls. Future coastal population growth and exposure to sea-level rise and coastal flooding - A global assessment. *PLOS ONE*, 10(3):1–34, 03 2015.

[84] Yoshimitsu Okada. Surface deformation due to shear and tensile faults in a half-space. *Bulletin of the Seismological Society of America*, 75(4):1135–1154, 1985.

[85] Peter R. Oke and Pavel Sakov. Representation error of oceanic observations for data assimilation. *Journal of Atmospheric and Oceanic Technology*, 25(6):1004–1017, 2008.

[86] Dinh Tuan Pham. Stochastic methods for sequential data assimilation in strongly nonlinear systems. *Monthly Weather Review*, 129(5):1194–1207, 2001.

[87] Dinh Tuan Pham, Jacques Verron, and Marie Christine Roubaud. A singular evolutive extended Kalman filter for data assimilation in oceanography. *Journal of Marine Systems*, 16(3):323 – 340, 1998.

[88] Chiara Piccolo and Mike Cullen. A new implementation of the adaptive mesh transform in the Met office 3D-Var system. *Quarterly Journal of the Royal Meteorological Society*, 138(667):1560–1570, 2012.

[89] P. Rebeschini and R. van Handel. Can local particle filters beat the curse of dimensionality? *ArXiv e-prints*, January 2013.

[90] A. R. Robinson, P. F. J. Lermusiaux, and N. Q. Sloan. Data assimilation. *The sea*, 10:541–594, 1998.

[91] Gerald L Smith, Stanley F Schmidt, and Leonard A McGee. *Application of statistical filter theory to the optimal estimation of position and velocity on board a circumlunar vehicle*. National Aeronautics and Space Administration, 1962.

[92] Chris Snyder and Fuqing Zhang. Assimilation of simulated Doppler radar observations with an Ensemble Kalman filter. *Monthly Weather Review*, 131(8):1663–1677, 2003.

[93] Costas E. Synolakis, Eddie N. Bernard, Vasily V. Titov, Utku Kanoglu, and Frank I. Gonzalez. Validation and verification of tsunami numerical models. *Pure and Applied Geophysics*, 165(11-12):2197–2228, 2008.

[94] Vasily V. Titov Costas Emmanuel Synolakis. Numerical modeling of tidal wave runup. *Journal of Waterway, Port, Coastal, and Ocean Engineering*, 124(4):157–171, Issue: object: doi:10.1061/jwped5.1998.124.issue-4, revision: rev:1479482265789-33025:doi:10.1061/jwped5.1998.124.issue-4, .

[95] L. Tang, V. V. Titov, and C. D. Chamberlin. Development, testing, and applications of site-specific tsunami inundation models for real-time forecasting. *Journal of Geophysical Research: Oceans*, 114(C12), 2009.

[96] Michael K Tippett, Jeffrey L Anderson, Craig H Bishop, Thomas M Hamill, and Jeffrey S Whitaker. Ensemble square root filters. *Monthly Weather Review*, 131(7):1485–1490, 2003.

[97] Vasily V Titov, Frank I Gonzalez, EN Bernard, Marie C. Eble, Harold O. Mofjeld, Jean C. Newman, and Angie J. Venturato. Real–time tsunami forecasting: Challenges and solutions. In *Developing tsunami-resilient communities*, pages 41–58. Springer, 2005.

[98] G. Triantafyllou, I. Hoteit, X. Luo, K. Tsiaras, and G. Petihakis. Assessing a robust ensemble-based Kalman filter for efficient ecosystem data assimilation of the Cretan Sea. *Journal of Marine Systems*, 125:90 – 100, 2013.

[99] G. Triantafyllou, I. Hoteit, and G. Petihakis. A singular evolutive interpolated Kalman filter for efficient data assimilation in a 3-d complex physical–biogeochemical model of the cretan sea. *Journal of Marine Systems*, 40–41:213 – 231, 2003.

[100] Rüdiger Verfürth. *A review of a posteriori error estimation and adaptive mesh-refinement techniques*. John Wiley & Sons Inc, 1996.

[101] Jérôme Vialard, Frédéric Vitart, Magdalena A. Balmaseda, Timothy N. Stockdale, and David L. T. Anderson. An ensemble generation method for seasonal forecasting with an ocean–atmosphere coupled model. *Monthly Weather Review*, 133(2):441–453, 2005.

[102] Xuguang Wang, Dale M. Barker, Chris Snyder, and Thomas M. Hamill. A Hybrid ETKF–3DVAR Data Assimilation Scheme for the WRF model. Part I: Observing system simulation experiment. *Monthly Weather Review*, 136(12):5116–5131, 2008.

[103] Yong Wei, Eddie N. Bernard, Liujuan Tang, Robert Weiss, Vasily V. Titov, Christopher Moore, Michael Spillane, Mike Hopkins, and Utku Kânoğlu. Real-time experimental forecast of the Peruvian tsunami of August 2007 for U.S. coastlines. *Geophysical Research Letters*, 35(4), 2008.

[104] Jeffrey S. Whitaker and Thomas M. Hamill. Ensemble data assimilation without perturbed observations. *Monthly Weather Review*, 130(7):1913–1924, 2002.

[105] Jeffrey S. Whitaker and Thomas M. Hamill. Evaluating methods to account for system errors in ensemble data assimilation. *Monthly Weather Review*, 140(9):3078–3089, 2012.

Vita

Pushkar Kumar Jain was born in Chennai, Tamil Nadu, India on 3 March 1991, the son of Rajendra Kumar Jain and Vasumati Jain. He received the Bachelor of Engineering (Hons.) in Mechanical Engineering from Birla Institute of Technology & Science - Pilani Hyderabad Campus. During the undergraduate degree, he worked as a research intern at the Council of Scientific & Industrial Research - National Aerospace Laboratories, Bangalore, India. He worked with the scientists towards development of composite materials for aerodynamic applications. Thereafter, he started working as a Consultant for General Electric - Global Research Center in Bangalore, India in 2013. As part of the Aero-thermal Mechanical Systems Laboratory, he was primarily involved in research and development of bearings for wind turbines. He applied to the University of Texas at Austin for enrollment in their Masters/PhD program of Engineering Mechanics. He was accepted and started graduate studies in August, 2013.

Permanent address: pushkarjain1991@gmail.com

This dissertation was typeset with LaTeX[†] by the author.

[†] LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.